



# Evaluation d'une interface par des ergonomes : diagnostics et strategies

Agnès Pollier

## ► To cite this version:

Agnès Pollier. Evaluation d'une interface par des ergonomes : diagnostics et strategies. [Rapport de recherche] RR-1391, INRIA. 1991. inria-00075170

**HAL Id: inria-00075170**

**<https://inria.hal.science/inria-00075170>**

Submitted on 24 May 2006

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNITÉ DE RECHERCHE  
IRIA-ROCQUENCOURT

# Rapports de Recherche

N° 1391

## *Programme 3*

*Intelligence artificielle, Systèmes cognitifs et  
Interaction homme-machine*

## **EVALUATION D'UNE INTERFACE PAR DES ERGONOMES : DIAGNOSTICS ET STRATEGIES**

**Agnès POLLIER**

Institut National  
de Recherche  
en Informatique  
et en Automatique

Domaine de Voluceau  
Rocquencourt  
B.P. 105  
78153 Le Chesnay Cedex  
France  
Tél. (1) 39 63 55 11

**Février 1991**



**Programme 3**

**EVALUATION D'UNE INTERFACE PAR DES ERGONOMES :  
DIAGNOSTICS ET STRATEGIES**

**User interface evaluation by human factors specialists :  
diagnoses and strategies**

**Agnès POLLIER**

**Janvier 1991**



## Résumé

Dans le contexte du diagnostic ergonomique d'interfaces, l'explicitation du savoir-faire des ergonomes constitue une source d'information très utile pour la définition d'une méthode d'évaluation et la spécification d'un environnement d'assistance.

Cette étude rend compte d'un premier examen de l'activité d'ergonomes en situation d'évaluation individuelle d'une interface existante (application de gestion de communications multi-média).

Les résultats concernent deux aspects de la performance d'évaluation : les diagnostics effectués et les stratégies mises en œuvre pour organiser l'examen de l'interface.

Concernant les diagnostics, nos résultats confirment ceux d'autres études : dans des conditions expérimentales restrictives, la performance individuelle d'évaluation est très variable, notamment quant aux types de problèmes évoqués. Cette étude apporte par ailleurs de nouveaux résultats sur les stratégies.

L'examen des stratégies montre que le processus d'évaluation est très complexe. Cinq stratégies très imbriquées, contribuent à des degrés et à des moments divers à la démarche d'évaluation : stratégies selon des objectifs d'utilisation, selon la structure de l'interface, selon des niveaux d'abstraction, selon les objets de l'interface et enfin selon des critères ergonomiques.

Les démarches des sujets varient selon les priorités d'examen adoptées au sein de ces stratégies, en particulier selon que leur approche est top-down ou bottom-up (stratégie selon les niveaux d'abstraction), en largeur d'abord ou en profondeur d'abord (stratégie selon la structure).

En plus d'indications méthodologiques intéressantes pour les futurs travaux, cette étude a permis de formuler un certain nombre d'enseignements concernant la définition de méthodes ou d'environnements d'aide à l'évaluation ergonomique d'interfaces.

**Mots-clés :** évaluation d'interfaces, expertise en ergonomie, stratégies, critères ergonomiques, méthode d'évaluation.

## Abstract

In the context of evaluating ergonomic aspects of computer interfaces, eliciting human factors expertise constitutes a useful source of information for defining evaluation methods and for specifying aiding tools.

This study presents an initial investigation of the activities of human factors professionals in the evaluation of an existing interface (a multi-media communication system).

The results concern two aspects of the evaluation behavior : the diagnosis made and the strategies used to organize the examination of the interface.

Concerning the diagnosis, the results confirm those of other studies : in restricted experimental conditions, individual evaluation performance is quite variable ; most notably in the types of problems identified. This study also provides new results about strategies.

The examination of the strategies show that the process of evaluation is very complex. Five overlapping strategies to some extent contribute to the evaluation and at different times. Strategies are based on : user's objectives ; interface structure ; various levels of abstraction ; interface objects and ergonomic criteria.

Within these strategies, the subjects approach varies according to priorities in the examination of the interface, particularly whether their approach is top-down or bottom up (strategies based on levels of abstraction), and whether their approach is depth first or breadth first examinations (strategies based on interface structure).

In addition to interesting methodological considerations for further work, this study provides a number of indications for defining methods or aiding environments for the ergonomic evaluation of interfaces.

**Key-words :** human factors evaluation, human factors expertise, strategies, ergonomics criteria, evaluation method.

## SOMMAIRE

1 INTRODUCTION.....	1
2 METHODE.....	6
2.1 L'interface .....	6
2.2 Les sujets.....	7
2.3 L'expérience.....	8
2.4 Données recueillies.....	9
2.5 Traitement des données.....	9
2.5.1 Classification des problèmes ergonomiques évoqués.....	9
2.5.2 Analyse des séquences d'évaluation et examen des régularités.....	10
3 LES PROBLEMES EVOQUES.....	10
3.1 Nombre et pourcentage de problèmes évoqués.....	11
3.2 Similarité des problèmes .....	12
3.2.1 Résultat global .....	12
3.2.2 Proximité entre les sujets.....	13
3.3 Types de problèmes.....	14
3.3.1 Affichage vs Séquences de commandes.....	14
3.3.2 Existence vs Inexistence .....	15
3.4 Critères d'évaluation, Standards et Problèmes Fonctionnels.....	17
3.4.1 Caractérisation des problèmes .....	17
3.4.2 Priorités d'évocation des critères.....	20
3.4.2.1 Hiérarchie globale des critères.....	20
3.4.2.2 Variabilité intersujets par critère.....	21
3.4.2.3 Hiérarchie des critères par sujet.....	22
3.4.3 Standards.....	24
3.4.4 Problèmes Fonctionnels.....	24
3.4.5 Interaction entre l'ensemble des descripteurs.....	26
3.5 Discussion.....	28
4 STRATEGIES D'EVALUATION.....	31
4.1 Méthodes d'analyses.....	31
4.1.1 Caractérisation des comportements élémentaires.....	32
4.1.2 Analyse des séquences d'évaluation minimales.....	32
4.1.3 Analyse de l'ensemble des séquences d'évaluation.....	33
4.2 Description des stratégies d'évaluation.....	34
4.2.1 Stratégie selon des objectifs d'utilisation .....	35
4.2.2 Stratégie selon la structure de l'interface.....	36
4.2.3 Stratégie selon les niveaux d'abstraction .....	37
4.2.4 Stratégie selon les objets de l'interface.....	39
4.2.5 Stratégie selon des critères.....	40
4.3 Mise en œuvre des stratégies.....	40
4.4 Discussion.....	45
5 CONCLUSION.....	48
REFERENCES.....	54
ANNEXES	

## 1 INTRODUCTION

Les méthodes d'évaluation ergonomique d'interfaces utilisateurs deviennent un enjeu majeur dans le domaine de la conception des logiciels interactifs.

Ces dernières années, de nombreuses recherches dans le domaine de l'interaction homme-ordinateur se sont centrées sur le processus de conception et sur la manière dont l'ergonomie y est abordée. Conscients du fait que la conception de logiciels adaptés à l'utilisateur est un problème particulièrement complexe et qu'il est impossible d'obtenir le "bon système" dès la première phase de conception, les chercheurs insistent sur la nécessité d'effectuer une conception itérative, avec plusieurs cycles de conception-évaluation et une participation précoce et continue des utilisateurs (Carroll et Rosson, 1985 ; Gould et Lewis, 1985). Des tests ergonomiques doivent être réalisés le plus tôt possible et pas seulement a posteriori lors d'une contribution évaluative extérieure (Scapin, 1986). L'évaluation ergonomique n'est donc plus considérée comme une étape spécifique se produisant à un moment particulier de la conception mais comme une activité continue tout au long du processus de conception.

Parallèlement, l'émergence d'outils de développement d'interfaces dans les ateliers de génie logiciels (boîtes à outils, gestionnaires de fenêtres, éditeurs d'interfaces) permet de réaliser rapidement des simulations de versions de logiciels non définitives (maquettes ou prototypes). Les concepteurs ont donc la possibilité de tester plus fréquemment et précocement leurs interfaces mais aussi de les modifier à moindre coût. L'évaluation et la modification appropriées de prototypes peuvent avoir des effets importants sur l'utilisation du produit final (Savage et al., 1982).

Cependant, le développement des outils de prototypage reste insuffisant pour garantir, d'une part, des évaluations plus fréquentes et plus précoces et, d'autre part, que les tests ergonomiques effectués soient appropriés. En effet, Rosson et al. (1988) montrent que la présence d'un environnement favorable à l'itération n'augmente pas la fréquence et la précocité des tests effectués, ni par ailleurs leur pertinence : dans la plupart des cas, l'évaluation est encore réalisée tardivement et plutôt de façon informelle, i.e. basée sur l'intuition et le bon sens des concepteurs.

Les outils de prototypage restent donc avant tout des moyens qui facilitent la mise en place d'évaluation d'interfaces, rapides et de moindre coût. Ils posent les mêmes problèmes de sélection concernant les objectifs, les dimensions et méthodes d'évaluation que des techniques plus anciennes du genre dessin d'écrans, spécifications papier, etc. Autrement dit, ils ne répondent pas aux questions : pourquoi on évalue, quels aspects évaluer et comment les évaluer ?

Quels sont donc les moyens dont disposent actuellement les concepteurs pour évaluer leurs interfaces ?

Différentes techniques sont disponibles pour évaluer les interfaces d'un point de vue ergonomique et peuvent être distinguées selon deux grands types d'approches (Senach, 1990) : une approche empirique et une approche analytique.

L'approche empirique de l'évaluation est basée sur le recueil et l'analyse de performances d'utilisateurs. Il s'agit de données comportementales (e.g. vitesse d'exécution ou d'apprentissage d'une tâche, jugements de satisfaction, verbalisations) qui rendent compte de l'utilisation d'une interface existante ou prototypée, en situation réelle ou en laboratoire. Les évaluations auprès d'utilisateurs peuvent être effectuées tout au long du processus de conception et peuvent servir à différents objectifs (e.g. tester des alternatives de conception, réaliser un diagnostic d'usage ou comparer deux interfaces).

Cependant, elles nécessitent généralement une formation ou des techniques relevant du domaine de l'ergonomie et de la psychologie (e.g. construction de questionnaires, analyse de protocoles verbaux) que n'ont pas nécessairement les concepteurs. Ainsi, un certain nombre de déficiences ont pu être relevées (Howard et Murray, 1987; Sweeney et Dillon, 1987) : absence ou manque d'explicitation claire des questions de départ, choix de techniques inadaptées entraînant le recueil d'une quantité ou d'un type d'information ne répondant pas aux questions posées, etc.

Par ailleurs, la mise au point d'une expérimentation contrôlée et l'enregistrement de données relatives à l'utilisation prennent du temps, ce qui s'avère souvent incompatible avec les exigences de rapidité de développement des systèmes.

De plus, les utilisateurs finaux ne sont pas toujours disponibles, et d'autre part, les concepteurs ne sont pas totalement convaincus semble-t-il, de l'utilité d'effectuer des tests précoces auprès d'utilisateurs. En effet, l'évaluation empirique est encore considérée beaucoup plus comme une source évaluative (i.e. ayant une fonction corrective) que comme une source génératrice d'information pour des solutions de conceptions (Rosson et al., 1988). Il reste donc à démontrer aux concepteurs, à travers des exemples concrets, l'impact que peuvent avoir des tests précoces auprès d'utilisateurs.

Etant donné que les évaluations empiriques semblent pour l'instant mal adaptées aux concepteurs ayant peu de formation en ergonomie, l'approche analytique peut paraître plus accessible aux concepteurs, même si dans certains cas des tests a posteriori seront toujours nécessaires. Basée sur des modèles théoriques de l'interaction homme-ordinateur, des recommandations ou des critères ergonomiques, l'évaluation analytique

offre la possibilité d'effectuer un diagnostic a priori des caractéristiques ergonomiques des interfaces, i.e. sans avoir recours à des données relatives à l'utilisation d'un système. En d'autres termes, elle cherche à prédire à partir des propriétés d'une interface, quelles seront les difficultés que les utilisateurs pourront rencontrer.

Au sein de l'approche analytique, on distingue différents types de contributions : les principes généraux, les modèles prédictifs de la performance d'utilisateurs, les modèles de la qualité de l'interface et enfin les recommandations ergonomiques.

- Les principes généraux concernent surtout des suggestions d'étapes pour la conception ou quelques recommandations importantes (e.g. commencer par définir les besoins des utilisateurs, l'utilisateur doit toujours avoir la possibilité de contrôler le dialogue, etc. (Norman, 1983)).

Cependant, quelques suggestions portent plus précisément sur le processus d'évaluation. Par exemple, Murray et Howard (1987) proposent un plan d'évaluation en 5 étapes : spécifier des objectifs d'évaluation, décider de l'information nécessaire pour atteindre ces objectifs, choisir des techniques d'évaluation capables de fournir cette information, recueillir et analyser les données, et enfin, faire des recommandations pour la reconception ou des évaluations ultérieures. Ces étapes restent très générales mais quelques auteurs (e.g. Howard et Murray, 1987; Sweeney et Dillon, 1987; Karat, 1988) tentent au travers d'exemples concrets d'évaluations, de souligner l'importance de certaines étapes ou de les définir plus précisément. Ainsi, il s'agit de déterminer, pour chacune des techniques d'évaluation disponibles, le contexte dans lequel elle semble le plus appropriée (i.e. pour recueillir tel type d'information, à tel moment du processus de conception, etc.).

- Les modèles prédictifs des performances d'utilisateurs, principalement les modèles de tâches (e.g. KLM ; GOMS (Card et al., 1983), les modèles linguistiques (e.g. ALG (Reisner, 1981) ; CLG (Moran, 1981)) et les modèles cognitifs (e.g. modèle de la complexité cognitive (Kieras et Polson, 1985)), permettent de prédire jusqu'à un certain point des performances d'utilisation (durées d'exécution ou occurrences d'erreurs).

Cependant, ces modèles sont actuellement encore insuffisamment développés ou validés pour constituer des outils d'évaluation à part entière. Par ailleurs, leur utilisation qui nécessite un travail de formalisation assez long est peu adaptée pour effectuer des tests en cours de conception. Enfin et surtout, ces modèles concernent uniquement certains aspects de l'interface (e.g. langage de commande) et sont utilisés essentiellement dans le cadre d'évaluations comparatives (Roberts et Moran, 1983).



- Les modèles de la qualité de l'interface cherchent à identifier les propriétés intrinsèques qui auront un effet sur la performance de l'utilisateur, notamment en limitant les difficultés d'utilisation. Dans cette perspective, il existe quelques travaux tentant d'évaluer la qualité ergonomique d'interfaces, sur certains aspects : par exemple, l'évaluation quantitative de la complexité perceptive des affichages (Streveler et Wasserman, 1985 ; Tullis, 1988). Là encore, les mesures proposées ne portent que sur des aspects très particuliers de l'interface.
- Enfin, les connaissances ergonomiques disponibles dans la littérature constituent une perspective intéressante pour effectuer un diagnostic a priori de la qualité d'une interface. En effet, il est possible à partir des recommandations, guides ou divers standards, de définir un ensemble de dimensions et propriétés permettant d'évaluer une interface (Cf. par exemple, la grille d'évaluation de Smith et Mosier (1984)).

Ce type d'approche pose encore un certain nombre de problèmes. En particulier, les recommandations sont souvent trop générales ou trop spécifiques pour être appliquées directement ; elles sont encore incomplètes, i.e. ne concernent pas tous les aspects d'une interface (e.g. il y a peu de recommandations sur les aspects sémantiques ou conceptuels), etc. Outre les problèmes liés à la nature même des recommandations, un des aspects les plus importants concerne leur organisation générale. En effet, lors d'une évaluation, par quel aspect ou quelle recommandation commencer et que faire ensuite ? Quelle recommandation choisir lorsque plusieurs sont applicables ? Un effort est en cours pour définir un cadre de travail permettant de décrypter et d'organiser de façon cohérente les diverses recommandations ergonomiques disponibles, notamment sous la forme de règles de production (Scapin, 1990a).

A l'issue de ce rapide examen des contributions destinées à évaluer la qualité ergonomique des interfaces utilisateurs, on constate qu'il existe des techniques variées mais encore insuffisantes pour constituer des méthodologies d'évaluation directes et facilement utilisables pour les concepteurs. Il s'avère donc nécessaire, pour aider les concepteurs peu familiers avec l'ergonomie à effectuer rapidement un bon diagnostic de leurs logiciels, de développer des méthodes, en particulier de type analytique.

Deux moyens sont envisageables :

- on peut tenter de définir "théoriquement" une méthode d'évaluation à partir d'une organisation des recommandations basée, par exemple, sur la structure des objets d'une interface, sur des critères ergonomiques ou sur des priorités entre recommandations ;
- on peut également envisager de dériver une méthode à partir de l'observation et de la formalisation de l'activité d'évaluation d'ergonomes expérimentés.

L'inconvénient de la première approche est qu'elle nécessite de faire un certain nombre de choix de façon arbitraire, dans la mesure où il est difficile, par exemple, d'établir a priori des priorités entre recommandations et d'identifier les solutions de compromis. De plus, selon cette approche, on ne peut garantir que le type de stratégie choisi puisse correspondre à ou s'insérer dans les stratégies habituelles des ergonomes.

Pour ces raisons, nous avons choisi de privilégier la seconde approche. Cela dit, l'intervention ergonomique étant actuellement, semble-t-il, plus basée sur une expertise que sur une méthodologie, il n'est pas garanti que les ergonomes eux-mêmes utilisent une méthode unique. Cependant, du fait qu'ils sont capables de détecter a priori (i.e. sans avoir recours à des performances d'utilisateurs) et rapidement les principaux défauts d'une interface, l'analyse de leur activité et en particulier de leurs stratégies d'évaluation peut constituer une source d'information précieuse concernant notamment :

- l'organisation générale de l'examen d'une interface. Par exemple, quels types d'objets ou de propriétés examiner en priorité, mais aussi comment prendre en compte des unités de plus haut niveau de l'interface (e.g. des transactions plus générales du dialogue) ?
- l'importance relative des problèmes ergonomiques rencontrés. En effet, il est raisonnable de penser que les ergonomes considèrent certains problèmes comme plus importants que d'autres, i.e. comme entraînant de plus grandes difficultés d'utilisation. Il sera donc intéressant de déterminer s'il existe une hiérarchie implicite ou explicite des problèmes détectés et sur quel(s) type(s) d'information(s) elle est basée (e.g. connaissances des caractéristiques de la tâche ou des utilisateurs, critères ergonomiques, etc.).

Par ailleurs, on peut également poser le problème d'un diagnostic global de la qualité d'une interface, i.e. comment les différents problèmes détectés peuvent-ils être traduits en un compte rendu plus général sur la facilité d'utilisation de l'interface ? Hammond et al.(1985) montrent à ce sujet que les experts, contrairement aux utilisateurs, sont capables d'intégrer un ensemble de problèmes particuliers en une appréciation globale de la qualité ergonomique d'un logiciel.

De manière générale on ne s'intéressera pas vraiment aux données ergonomiques individuelles évoquées par les ergonomes, mais plutôt à leur organisation. En effet, les données ergonomiques de base sont par ailleurs considérées à partir de l'ensemble des résultats de la littérature (Aschehoug, 1989 ; Scapin, 1990a).

L'étude<sup>1</sup> présentée dans ce rapport rend compte d'un premier examen de l'activité de quatre ergonomes expérimentés en situation d'évaluation d'une interface existante.

On s'est intéressé principalement à deux aspects de cette activité :

- d'une part, aux résultats des diagnostics effectués : quels sont les problèmes ergonomiques détectés, selon quels critères, selon quelles priorités ? Mais aussi, dans quelle mesure les diagnostics des différents experts se recouvrent-ils ?
- d'autre part, aux stratégies mises-en-œuvre pour examiner et évaluer l'interface, i.e. comment l'examen des divers aspects de l'interface est-il organisé, selon quelles dimensions, avec quelle priorité ? Observe-t-on une approche systématique dans la démarche d'évaluation concernant par exemple, les différents objets de l'interface et/ou la prise en compte de certains critères ?

## 2 METHODE

### 2.1 L'interface

L'étude a été conduite sur l'interface d'une application de gestion de communications multi-média fonctionnant sur un poste PC et programmée sous l'environnement MS-Windows. Cette application a été développée par une société de services<sup>2</sup> pour le CNET<sup>3</sup>. Elle a été conçue pour exploiter les diverses opportunités offertes par la connexion d'un ou de plusieurs micro-ordinateurs à un nouveau type de réseau de télécommunications, le RNIS<sup>4</sup>. Ce dernier permet de faire transiter sur une liaison unique, à très grande vitesse et simultanément, des informations d'origines diverses (voix, données informatiques, images).

L'application de gestion de communications permet donc aux utilisateurs de micro-ordinateurs reliés au réseau d'établir simultanément plusieurs types de communications (téléphone, transfert de fichiers, télétype, vidéotex (minitel), etc.) mais aussi de disposer de nouveaux services. Par exemple, l'utilisateur peut connaître l'identité de son correspondant avant de prendre un appel ou encore a la possibilité de laisser un message écrit sur le poste d'un correspondant absent ou occupé.

---

<sup>1</sup> Cette étude a été réalisée dans le cadre d'une convention de recherche entre l'INRIA et le CNET (n° 88 7B 016 00 790 92 45 LAA/TSS/SMD).

<sup>2</sup> La société ARIES.

<sup>3</sup> Centre National d'Etudes pour les Télécommunications.

<sup>4</sup> Réseau Numérique à Intégration de Services (Produit CNET).

Au moment de l'expérience, le logiciel fonctionnait encore à l'état de prototype et une partie seulement des fonctionnalités prévues étaient en service. Un des projets du CNET était justement d'évaluer ce premier prototype, d'en améliorer l'interface et d'étendre les fonctionnalités existantes.

## 2.2 Les sujets

Quatre ingénieurs ergonomes expérimentés (entre cinq et douze ans d'expérience en ergonomie de l'informatique) ont accepté d'être sujets de l'expérience.

Tous ont déjà participé à la conception - évaluation d'interfaces et de dialogues développés dans des domaines d'application divers (minitel, téléphonie, systèmes audiovisuels domestiques, éditeurs graphiques, etc.).

Néanmoins, leur expérience en matière d'évaluation est variable : deux d'entre eux sont intervenus à plusieurs reprises à titre de consultants pour effectuer des diagnostics d'interfaces ; les deux autres ont principalement travaillé dans la recherche de solutions de conception.

Enfin, concernant le produit à évaluer, nous avons relevé différents éléments qui pouvaient avoir une influence sur la performance d'évaluation : la connaissance de l'application, de la composante interface-système MS-Windows et du réseau RNIS (Cf. tableau 1).

Tableau 1  
Connaissances préalables des sujets concernant le produit à évaluer

Sujets	Connaissances sur le produit		
	Application	MS-Windows	RNIS
Sujet 1	un peu <sup>1</sup>	oui	oui
Sujet 2	non	non	oui
Sujet 3	non	oui	un peu
Sujet 4	non	oui	un peu

<sup>1</sup> Il s'agit de l'ergonome chargé de reconcevoir l'interface. Deux mois avant l'expérience il avait réalisé une exploration rapide de cette dernière mais il ne connaissait pas toutes les fonctionnalités disponibles, ni les modifications qui avaient pu être effectuées depuis.

## 2.3 L'expérience

L'expérience s'est réalisée individuellement et comporte deux phases :

Première phase : présentation de l'application et contrôle des connaissances

L'expérimentateur présente oralement et brièvement l'application à évaluer à chacun des sujets comme ci-dessous : "L'interface que vous aurez à évaluer est une application de gestion de communication fonctionnant sur PC dans l'environnement MS-Windows. Elle permet d'effectuer des communications téléphoniques ou des transmissions de données comme par exemple du transfert de fichiers ou de textes. Il y a la possibilité d'établir simultanément les deux types de communications, téléphone et transfert de données".

Il procède ensuite à un entretien dirigé afin de noter :

- les connaissances préalables que possède l'ergonome sur l'application à évaluer (concernant le RNIS, l'environnement MS-Windows, et l'application ARIES)
- le niveau d'expertise en ergonomie : nombre d'années d'expérience professionnelle ; type d'activité réalisé (conception et/ou évaluation) ; domaine d'application (informatique, téléphonie, etc.) ; et formation universitaire.

Deuxième phase : évaluation de l'interface

Chaque sujet est placé devant un PC et a pour tâche d'évaluer l'application gestion des communications. L'évaluation ne doit pas dépasser deux heures. Les sujets doivent penser à voix haute et expliciter leur raisonnement ergonomique.

Outre l'expérimentateur, un expert qui connaît bien l'application est à la disposition des sujets pour répondre à leurs questions, leur fournir des informations et/ou explications concernant l'application (technique de l'information à la demande). Il peut également faire une démonstration si ceux-ci le souhaitent.

Cette situation d'étude a semblé intéressante pour recueillir des données "réalistes" sur les stratégies d'évaluation des ergonomes. En effet, d'une part, il s'agit d'un problème d'évaluation réel, non construit (en particulier non simplifié) pour les besoins de l'expérience et d'autre part, assez représentatif du type de tâche auquel les ergonomes peuvent être confrontés dans le cadre de leur travail. Il est en effet souvent demandé à un ergonome d'évaluer une interface qu'il ne connaît pas ou très peu afin d'obtenir un avis sur sa qualité ergonomique. Par ailleurs, la méthode de verbalisation simultanée à l'exécution de la tâche reste plus fiable que la technique de l'interview pour recueillir des données sur la façon dont les ergonomes procèdent réellement pour évaluer des

interfaces. En effet, en demandant simplement à des ergonomes quelles stratégies ou méthodes ils adoptent pour évaluer une interface, on s'expose au fait qu'ils structurent et rationalisent leur démarche beaucoup plus qu'elle ne l'est en réalité.

## **2.4 Données recueillies**

Chacune des passations de l'expérience a été filmée sur bande vidéo-son. Ceci a permis de disposer des productions verbales des sujets, synchronisées avec les états des écrans et du dialogue ainsi que des traces des interactions sur le poste au cours de l'évaluation.

De plus, on a recueilli les productions écrites éventuelles (dessins, schémas, etc.) ainsi que les autres sources d'informations utilisées (documents et autres outils).

L'analyse des bandes vidéo a consisté à recueillir différents paramètres :

- les problèmes ergonomiques évoqués, leur ordre et leur contexte d'évocation ;
- les différents états de l'interface explorés (chaque état étant caractérisé par un écran ou une fenêtre particulière) et l'ordre d'exploration ;
- les différents aspects examinés et/ou prises d'informations effectuées au sein de ces états.

Le recueil de ces paramètres a permis de constituer un protocole pour chaque sujet à partir duquel différents types d'analyses ont pu être effectués.

## **2.5 Traitement des données**

Le traitement des données s'est effectué selon deux axes correspondant aux deux principaux objectifs de cette étude, à savoir : d'une part, caractériser et comparer les diagnostics effectués et d'autre part, identifier les stratégies d'examen des divers aspects de l'interface.

### **2.5.1 Classification des problèmes ergonomiques évoqués**

Dans un premier temps, les problèmes ergonomiques détectés ont été classés pour établir une structure typologique qui permette non seulement de regrouper des problèmes similaires dans une même classe mais aussi de disposer d'un ensemble de descripteurs de la performance d'évaluation des sujets.

Différents types d'attributs ont été retenus pour caractériser les problèmes évoqués : critère ergonomique sous-jacent ; aspect et objet de l'interface concernés (e.g. affichage, entrée ou séquence de commandes) ; nature du problème (e.g. affichage inadéquat, manquant ou inutile) ; lieu (état ou fenêtre) et ordre d'évocation.

Quelques exemples d'états de l'interface sont présentés avec leurs codages respectifs en annexe 1. La structure typologique résumant les attributs utilisés et un exemple de classification des problèmes évoqués sont présentés en annexe 2.

### 2.5.2 Analyse des séquences d'évaluation et examen des régularités

Pour identifier les stratégies mises en œuvre par les ergonomes au cours de l'évaluation, on a recherché des régularités comportementales au sein des protocoles. Pour cela, on a examiné la similitude des comportements des sujets en augmentant progressivement la taille des séquences d'évaluation prises en compte. Trois types d'analyses ont été effectués : caractérisation des comportements élémentaires, analyse des séquences d'évaluation minimales et analyse de l'ensemble des séquences d'évaluation. Ces analyses seront décrites lorsque l'on abordera les résultats concernant les stratégies.

La performance d'évaluation des sujets va être présentée successivement selon deux perspectives : les problèmes évoqués et les stratégies mises en œuvre. Il s'agit respectivement de dénombrer et de caractériser les problèmes évoqués, et de tirer de l'examen des séquences d'évaluation des considérations sur les stratégies.

## 3 LES PROBLEMES EVOQUES

En premier lieu, on constate une relative homogénéité entre les sujets du point de vue des durées de passation. Alors que la consigne de l'expérience était de ne pas dépasser deux heures, tous les sujets ont terminé leur évaluation au bout d'une heure 15 minutes environ. Les différences éventuelles de performance entre sujets ne pourront donc pas être attribuées au fait que certains aient pris beaucoup plus de temps que d'autres.

Différents types d'indicateurs ont été retenus pour caractériser et comparer la performance des sujets du point de vue des diagnostics effectués : nombre, pourcentage, similarité et types de problèmes évoqués, critères d'évaluation et autres descripteurs, interactions entre descripteurs. On présente ci-dessous les résultats obtenus avec chacun de ces indicateurs.

### 3.1 Nombre et pourcentage de problèmes évoqués

Le nombre de problèmes diagnostiqués par chaque sujet est important : il varie entre 63 et 84 problèmes par sujet (avec une moyenne de 70 problèmes par sujet).

Le sujet 1 a évoqué un nombre relativement plus important de problèmes que les trois autres (Cf. tableau 2). Ceci peut s'expliquer par une meilleure connaissance préalable de l'application chez ce sujet mais aussi par le fait qu'il était plus directement impliqué dans le projet de reconception de cette interface.

Pour avoir une idée de l'importance du nombre de problèmes évoqués par chaque sujet, il aurait fallu pouvoir déterminer le nombre total de problèmes effectivement présents au sein de l'interface (de façon empirique ou à l'aide d'autres experts), ce que nous n'avons pu faire compte tenu des contraintes d'accès au système.

Cependant, on peut faire l'hypothèse qu'en rassemblant les problèmes distincts évoqués par nos quatre sujets, on obtient un nombre total de problèmes assez représentatif de l'ensemble des problèmes potentiellement présents dans cette interface. Ainsi, sur un total de 280 énoncés de problèmes, on a répertorié 167 problèmes différents, i.e. distincts. C'est sur cette base que l'on a calculé les pourcentages de problèmes évoqués par chaque sujet mais aussi par deux et trois sujets. Les résultats sont présentés dans le tableau 2.

Tableau 2  
Nombre et pourcentage de problèmes distincts  
évoqués respectivement par un, deux et trois sujets.

Sujets	Nombre de problèmes distincts évoqués	% de problèmes distincts évoqués (/167)	% moyen de problèmes distincts évoqués
Sujet 1	84	50	42
Sujet 2	69	41	
Sujet 3	64	38	
Sujet 4	63	38	
Sujets 1 et 2	124	74	67
Sujets 1 et 3	113	68	
Sujets 1 et 4	119	71	
Sujets 2 et 3	109	65	
Sujets 2 et 4	108	65	
Sujets 3 et 4	100	60	
Sujets 1, 2 et 3	146	87	85
Sujets 1, 2 et 4	150	89	
Sujets 1, 3 et 4	137	82	
Sujets 2, 3 et 4	137	82	



En considérant, dans le meilleur des cas, que les quatre sujets ont repéré la quasi-totalité des problèmes de cette interface, on constate qu'un sujet trouve en moyenne 42% des problèmes et au mieux la moitié des problèmes (sujet 1).

Le regroupement des diagnostics de deux sujets permet d'obtenir en moyenne 67% des problèmes. Autrement dit, il faut réunir les diagnostics d'au moins trois des sujets pour obtenir une image complète de l'ensemble des problèmes de cette interface (plus de 80% des problèmes).

### 3.2 Similarité des problèmes

#### 3.2.1 Résultat global

Il s'agit ici de déterminer plus précisément le degré de similarité des problèmes évoqués par les sujets. Pour cela, on a classé l'ensemble des problèmes selon qu'ils ont été évoqués par tous les sujets, par trois sujets, par deux ou encore par seulement un seul des sujets.

Deux problèmes ont été jugés équivalents lorsque leurs énoncés étaient similaires et/ou lorsqu'ils reposaient sur les mêmes observations (e.g. "l'icône S est affichée en permanence mais elle est inutile et inintéressante pour l'utilisateur" ; "l'icône S est inutile, elle ne sert à rien" ; "il n'y a pas de raison qu'il y ait cette icône qui apparaisse avec rien derrière"). Dans la plupart des cas, la caractérisation de similarité des problèmes a pu se faire de manière non ambiguë. Cependant, dans quelques cas, les décisions d'affectation ont été plus difficiles en raison des variations dans la généralité de l'expression des sujets.

Une comparaison du nombre de problèmes évoqués en commun par les quatre sujets par rapport au nombre total de problèmes distincts évoqués (167) a permis d'établir la proportion (%) de problèmes communs évoqués. De façon similaire, on a calculé la proportion de problèmes évoqués par trois sujets (trio), deux sujets (duo) et un seul sujet (spécifique). Le détail des calculs figure en annexe 3.

Les résultats obtenus sont résumés dans le tableau 3.

Tableau 3  
Pourcentages de problèmes évoqués selon leur degré de communauté

SPECIFIQUE	DUO	TRIO	COMMUN	TOTAL
59	21	14	6	100
59		41		100

Ce tableau met en évidence un très faible pourcentage de problèmes communs à l'ensemble des sujets (6%) et à l'inverse, un fort pourcentage de problèmes spécifiques (59%). Il reste que 41% des problèmes ont été évoqués par au moins deux sujets, dont 21% évoqués par deux sujets, 14 % par trois et 6% par quatre.

### 3.2.2 Proximité entre les sujets

De façon plus précise, on s'est intéressé à la proximité qui pouvait exister entre les sujets du point de vue des problèmes évoqués conjointement. Pour cela, on a examiné les pourcentages de problèmes communs chez les sujets pris deux à deux.

Les résultats obtenus sont présentés dans la figure 1. Pour chaque paire de sujets est représentée la part de problèmes communs par rapport à celle des problèmes spécifiques de chacun des sujets. Les pourcentages représentent la proportion de problèmes communs vis-à-vis de l'ensemble des problèmes distincts évoqués par chacune des paires de sujets.

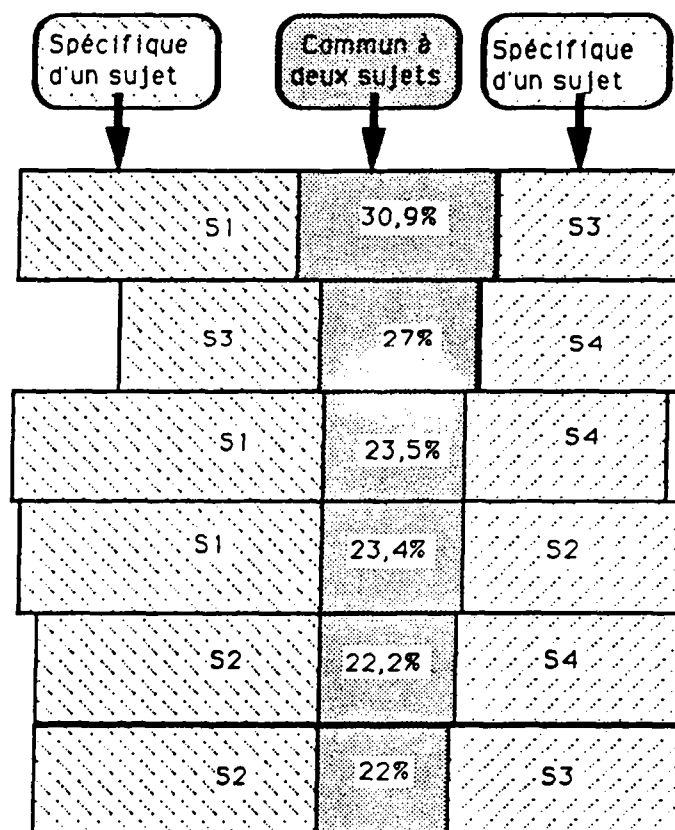


Fig.1 - Proportions relatives de communauté de problèmes évoqués par paire de sujets

On constate ainsi que la part de problèmes communs à deux sujets ne dépasse pas 31% (en moyenne 25%) et est plus ou moins importante selon les paires de sujets.

Les paires de sujets 1 et 3 ainsi que 3 et 4 ont, de ce point de vue, des pourcentages de communauté de problèmes importants, la communauté de problèmes entre les sujets 1 et

4 étant plus faible. Les paires de sujets incluant le sujet 2 ont quant à elles une communauté de problèmes plus faible.

Il conviendra d'examiner dans quelle mesure la part de problèmes communs à deux sujets peut être mise en relation avec une relative proximité de ces sujets du point de vue des critères, objectifs ou stratégies d'évaluation. Il ne semble pas en tout cas que le facteur "expérience professionnelle" (nombre d'années, domaine d'application ou type d'activité) permette d'apporter une explication satisfaisante à cette plus ou moins grande proximité entre les sujets.

### 3.3 Types de problèmes

L'objectif est de caractériser et de comparer les diagnostics effectués du point de vue de la complexité du raisonnement mis en œuvre. En d'autres termes, on fait l'hypothèse que la détection de certains types de problèmes nécessite un examen de l'interface plus complexe et plus élaboré que d'autres.

Pour tester le bien fondé de cette hypothèse, on se propose de comparer successivement les pourcentages de problèmes évoqués par les sujets sur deux dimensions :

- problèmes d'affichage vs problèmes concernant les séquences de commandes du système ;
- problèmes "d'existence" (i.e. diagnostics d'amélioration) vs problèmes "d'inexistence" (i.e. diagnostics d'absence).

#### 3.3.1 Affichage vs Séquences de commandes

L'hypothèse sous-jacente à cette comparaison est la suivante :

- les problèmes d'affichage concernent plutôt des aspects de surface de l'interface, i.e. des aspects qui sont immédiatement accessibles au raisonnement ;
- les problèmes concernant les séquences de commandes portent sur des aspects moins directement accessibles au raisonnement. Ils nécessitent de prendre en compte des transactions plus importantes et impliquent donc la mise en œuvre d'un examen plus élaboré de l'interface.

Les pourcentages de problèmes appartenant respectivement aux catégories Affichage et Séquences de commandes sont présentés, par sujet, dans le tableau 1 en annexe 4. Ils

ont été calculés par rapport à l'ensemble des problèmes évoqués par chaque sujet<sup>1</sup>.

L'histogramme de la figure 2 permet d'illustrer pour chaque sujet, la part relative de problèmes appartenant à chacune de ces catégories.

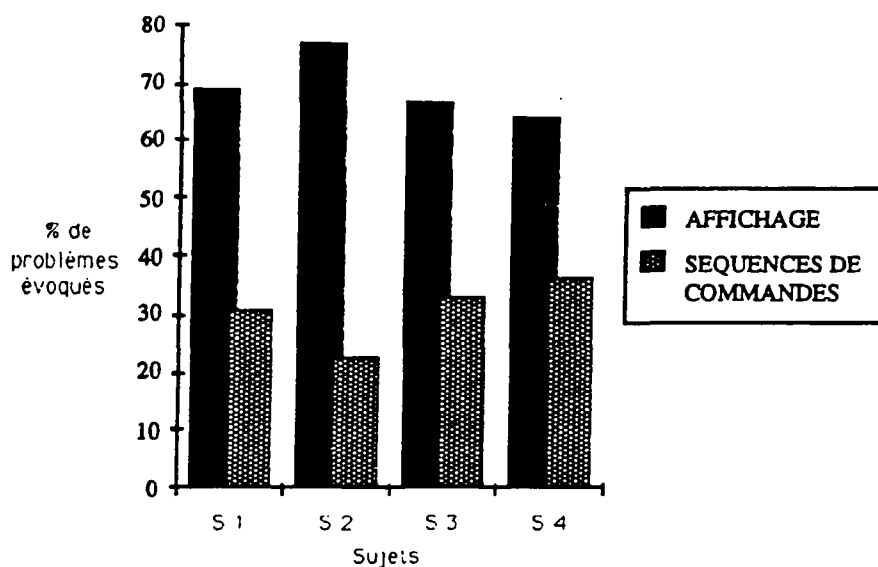


Fig.2 - Part relative des problèmes évoqués appartenant aux catégories Affichage et Séquences de commandes

Tous les sujets ont évoqué un plus grand pourcentage de problèmes (au moins deux fois plus) appartenant à la catégorie Affichage. Il existe donc une relative homogénéité entre les sujets.

Cependant, le sujet 2 évoque proportionnellement une plus grande part de problèmes (trois fois plus) appartenant à la catégorie Affichage que les trois autres sujets. Selon l'hypothèse de départ, cela signifierait chez ce dernier un degré moindre d'élaboration dans l'examen de l'interface. Toutefois, comme ce sujet était peu habitué à manipuler ce type d'interface, les aspects apprentissage ont pu interférer avec les aspects strictement évaluatifs.

### 3.3.2 Existence vs Inexistence

Il s'agit ici d'effectuer une distinction entre :

- d'une part, les "problèmes d'existence" : ils reposent sur la remise en cause d'éléments existants de l'interface, jugés inadéquats ou inutiles et suggèrent par conséquent des modifications de ces éléments (diagnostic d'amélioration) ;

<sup>1</sup> Les problèmes fonctionnels n'ont pas été inclus dans les calculs "Types de Problèmes" parce qu'ils n'ont pas été classés en fonction de ces dimensions.

- et, d'autre part les "problèmes d'inexistence" : ils dénoncent l'absence d'éléments et suggèrent donc l'introduction de nouveaux éléments (diagnostic d'absence).

L'hypothèse sous-jacente est que, comme les problèmes d'existence reposent sur l'examen d'éléments observables, ils sont plus immédiats et nécessitent un travail d'élaboration moins important que les problèmes d'inexistence.

Les pourcentages de problèmes appartenant respectivement aux catégories Existence et Inexistence sont présentés, par sujet, dans le tableau 2 en annexe 4. Ils ont été calculés de la même façon qu'avec l'indicateur précédent.

Les résultats obtenus sont illustrés à travers l'histogramme de la figure 3.

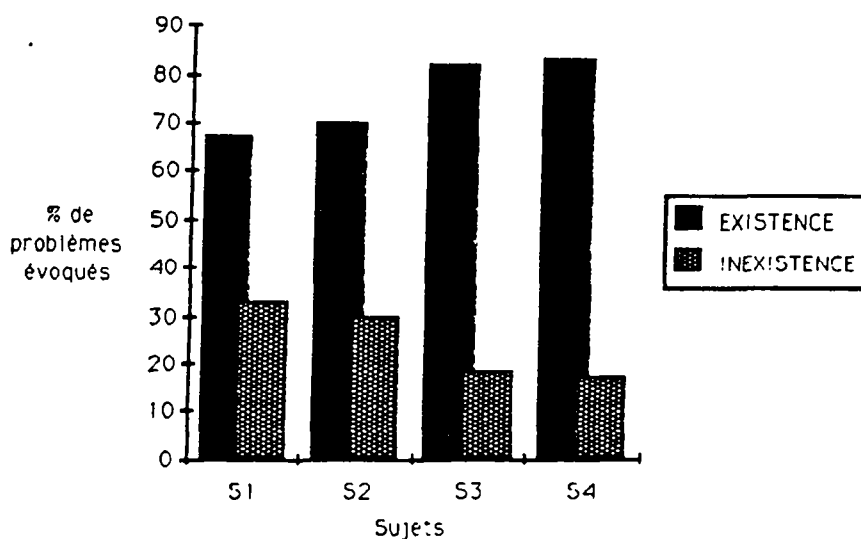


Fig.3 - Part relative des problèmes évoqués appartenant aux catégories Existence et Inexistence

On constate que tous les sujets évoquent une plus grande proportion de problèmes liés à la remise en cause d'éléments existants que de problèmes liés à l'absence d'éléments.

Cependant, les sujets 1 et 2 évoquent proportionnellement une plus grande part de problèmes d'inexistence que les sujets 3 et 4.

Une explication possible de ces résultats serait la connaissance plus importante des sujets 1 et 2 concernant le type de tâche à évaluer. En effet, les domaines d'application dans lesquels ils ont travaillé (téléphonie, minitel, audioconférence) autorisent un transfert plus important vers les notions de gestion des communications.

En confrontant ces résultats avec ceux obtenus lors de la comparaison Affichage vs Séquences de commandes, on constate qu'il n'y a pas de relation entre ces deux

indicateurs.

En particulier, le sujet 2 qui a été distingué précédemment par une proportion moins importante de problèmes liés aux séquences de commandes est ici caractérisé par une forte proportion de problèmes d'inexistence, ce qui paraît contradictoire du point de vue de nos hypothèses. La même remarque peut être faite pour les sujets 3 et 4 qui obtiennent des résultats opposés selon ces deux dimensions (proportion importante de problèmes liés aux séquences de commandes et faible proportion de problèmes d'inexistence). En conséquence, ces indicateurs ne semblent pas être pertinents en terme d'examen plus ou moins complexe ou élaboré de l'interface.

Par ailleurs, il est aussi important de noter que les problèmes fonctionnels n'ont pas été pris en compte lors de ces comparaisons puisqu'ils s'insèrent mal dans une structure du type Affichage vs Séquences de commande et/ou Existence vs Inexistence. Il sera intéressant de voir dans quelle mesure la comparaison des sujets du point de vue des problèmes fonctionnels évoqués rejoint les résultats obtenus ci-dessus.

Il faudra également examiner si les stratégies des sujets permettent d'explorer plus en avant cette question de niveau d'élaboration et/ou d'abstraction.

### 3.4 Critères d'évaluation, Standards et Problèmes Fonctionnels

#### 3.4.1 Caractérisation des problèmes

Deux types de descripteurs, des critères ergonomiques d'une part, les standards et les problèmes fonctionnels d'autre part, ont permis de caractériser les problèmes évoqués.

Les énoncés des problèmes ont été tout d'abord confrontés aux définitions d'un jeu de huit critères ergonomiques, établies à partir de la littérature et de l'examen de recommandations ergonomiques (Aschehoug, 1989 ; Scapin, 1990b).

Les critères sont rapidement définis ci-dessous<sup>1</sup> :

- Charge de Travail (CHTR)<sup>2</sup> :

Ce critère concerne l'ensemble des éléments de l'interface qui ont un rôle pour l'utilisateur, dans la réduction de sa charge perceptive ou mnésique et dans l'augmentation de l'efficacité du dialogue. Il comprend deux sous-critères : Brièveté et Charge Mentale.

- Compatibilité (COMP) :

Ce critère se réfère à l'accord pouvant exister entre les caractéristiques de l'utilisateur (perceptions, mémoires, habitudes, etc.) et l'organisation des entrées, des sorties et du dialogue.

<sup>1</sup> Le lecteur pourra trouver une définition plus détaillée de ces critères dans Aschehoug (1989) et Scapin (1990b)

<sup>2</sup> Les abréviations entre parenthèses correspondent au codage adopté pour l'affectation des critères aux problèmes.

- "Consistence" (CONS) :

Ce critère se réfère à la façon dont on effectue des choix similaires d'interface (codes, procédures, dénominations, etc.) dans des contextes identiques. Il s'applique à la localisation des objets, à leurs formats et leurs dénominations mais aussi à la syntaxe des commandes.

- Contrôle Utilisateur (COUT) :

Ce critère se réfère au contrôle dont dispose l'utilisateur sur le dialogue ainsi qu'au caractère explicite de ses actions.

- Adaptabilité (ADAP) :

Ce critère se réfère à la capacité d'un système à réagir différemment selon le contexte, les besoins et les préférences de l'utilisateur. Il comprend deux sous-critères : Flexibilité et Prise en compte de l'expérience de l'utilisateur.

- Gestion des Erreurs (GEER) :

Ce critère concerne tous les moyens mis en œuvre pour éviter et réduire les erreurs mais aussi pour les corriger. Il comprend trois sous-critères : Protection contre les erreurs, Qualité des messages d'erreurs et Correction des erreurs.

- Guidage (GUID) :

Ce critère concerne l'ensemble des moyens mis en œuvre pour conseiller, orienter, informer, et conduire l'utilisateur lors de ses interactions avec le système. Il comprend quatre sous-critères : Prompting, Distinction vs Regroupement entre items, Feed-back et Clarté.

- Signifiante des Codes (SICO) :

Ce critère se réfère à une bonne définition, i.e. claire et univoque des divers codes d'une interface : titres d'écran, labels des commandes ou des options, etc.

La mise en correspondance des arguments fournis spontanément par les sujets avec les définitions de ces critères a permis d'attribuer un (ou plusieurs) critère(s) à la plupart des problèmes.

Bien évidemment, les sujets n'ont pas toujours adopté spontanément les mêmes dénominations des critères que celles utilisées dans cette étude. En effet, un certain nombre de synonymes ou de périphrases ont été utilisés pour désigner spontanément ces critères. Ainsi, on a relevé l'emploi de la notion de cohérence pour désigner les problèmes d'homogénéité et de consistance ; celle de transparence pour désigner des problèmes de guidage ; celle de correspondance avec la logique de l'utilisateur pour des problèmes de compatibilité.

En plus des critères précédents, deux autres descripteurs ont été utilisés :

- les Standards (STAN) : il s'agit des problèmes caractérisés spontanément par les ergonomes comme traduisant le non respect de standards en particulier des standards MS-Windows ;
- les Problèmes Fonctionnels (PRFO) : il s'agit des problèmes qui reposent sur une remise en cause des fonctionnalités existantes et dont on peut penser qu'ils sont en relation assez directe avec une prise en compte de la tâche utilisateur (e.g. les fonctionnalités existantes permettent-elles à l'utilisateur d'atteindre ses objectifs ?).

Les problèmes fonctionnels concernent plutôt l' "utilité" de l'interface et donc surtout ses aspects conceptuels et sémantiques (capacité fonctionnelle, contenu informationnel de l'interface).

Par opposition, les critères et standards permettent plutôt de caractériser les problèmes ergonomiques ayant une influence sur l' "utilisabilité" de l'interface, i.e. sur sa facilité d'utilisation et d'apprentissage. Ils concernent donc les aspects de l'interface de plus bas niveaux d'abstraction que les problèmes fonctionnels, i.e. plutôt les aspects syntaxiques et lexicaux.

De façon générale, la classification des problèmes en fonction des critères, standards et problèmes fonctionnels a pu se faire sans trop d'ambiguïté. Néanmoins, dans certains cas et en particulier lorsque les arguments évoqués ne se référaient pas de façon explicite à un critère particulier, l'affectation a été plus difficile. Par exemple, un nombre de couleurs trop important au sein d'un même écran peut traduire un problème de charge mentale et/ou un problème de clarté. Dans la mesure où le sujet n'apporte pas d'explication supplémentaire, il est difficile de savoir à quel(s) critère(s) il se réfère, i.e. s'il se réfère à l'un ou à l'autre ou encore aux deux en même temps.

Par ailleurs, ce n'est pas parce qu'un sujet fait référence explicitement à un seul critère qu'il ne pense pas éventuellement à un autre. Pour ces raisons, on a extrait dans quelques cas, un ou plusieurs critères supplémentaires. Toutefois, ces critères n'ont pas été pris en compte au niveau du comptage des critères évoqués.

En outre, cette classification a fait l'objet d'une validation auprès d'une deuxième personne experte en ergonomie (accord entre deux juges).

Les pourcentages de problèmes affectés respectivement aux critères, standards et problèmes fonctionnels sont présentés, globalement (tous sujets) et par sujet, dans le tableau 4.

Tableau 4  
Pourcentages de problèmes évoqués affectés respectivement  
aux Critères, Standards et Problèmes Fonctionnels

Descripteurs	Sujets				
	Tous sujets	Sujet 1	Sujet 2	Sujet 3	Sujet 4
Critères	83	79	92	74	86
Standards et Problèmes Fonctionnels	17	21	8	26	14



La mise en correspondance des énoncés de problèmes avec les définitions des critères ergonomiques établies à partir de l'examen des recommandations a donc permis de caractériser 83% des problèmes évoqués. Ces critères d'évaluation semblent donc assez représentatifs des critères utilisés par les ergonomes. En tout état de cause, ils permettent une description organisée des problèmes.

Les standards (6%) et les problèmes fonctionnels (11%) représentent 17% des problèmes évoqués.

On constate par ailleurs que la part de problèmes affectés aux critères est plus ou moins importante selon les sujets.

Dans un premier temps, on s'intéressera aux priorités d'évocation des critères, globalement et par sujet. La variabilité intersujet du point de vue de la part de problèmes affectés respectivement aux catégories standards et problèmes fonctionnels sera ensuite examinée. Enfin, on procèdera à l'examen des interactions entre descripteurs évoqués simultanément.

### 3.4.2 Priorités d'évocation des critères

A partir de l'ensemble des problèmes évoqués affectés aux différents critères, trois types d'examen ont été effectués :

- hiérarchie globale des critères, i.e. priorité d'évocation des critères pour l'ensemble des sujets ;
- variabilité intersujets par critère ;
- hiérarchie des critères par sujet, i.e., variabilité intersujet du point de vue des priorités d'évocation des critères.

#### 3.4.2.1 Hiérarchie globale des critères

L'objectif est ici de déterminer pour l'ensemble des sujets si certains critères ont été évoqués plus souvent que d'autres, i.e. d'examiner la présence d'une éventuelle priorité ou hiérarchie des critères.

Les pourcentages d'évocation des différents critères par l'ensemble des sujets (nombre de problèmes évoqués pour un critère / nombre total de problèmes affectés aux différents critères) sont présentés dans le tableau 1 en annexe 5.

L'histogramme de la figure 4 illustre les résultats obtenus ; les critères ont été représentés par ordre d'évocation décroissant.

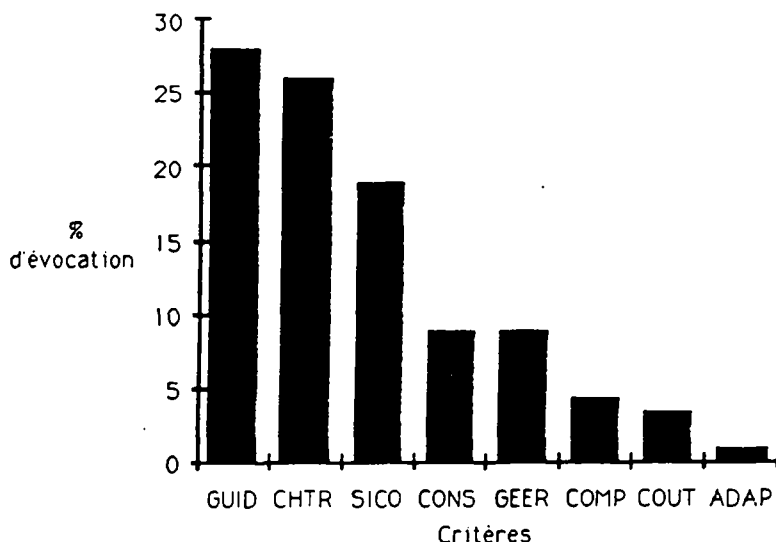


Fig. 4 - Hiérarchie globale des critères

On constate que ce sont les critères Guidage (GUID) et Charge de Travail (CHTR) qui ont été évoqués le plus grand nombre de fois. Le critère Signifiante des Codes (SICO) est également assez important. Ces trois critères représentent près de 70% des critères évoqués. Par contre les critères Compatibilité (COMP), Contrôle Utilisateur (COUT) et Adaptabilité (ADAP) sont très peu évoqués (moins de 5%) ; les critères Gestion des Erreurs (GEER) et Consistance (CONS) constituent des cas intermédiaires (entre 5 et 10%).

#### 3.4.2.2 Variabilité intersujets par critère

Les pourcentages de problèmes évoqués par critère et par sujet sont présentés dans le tableau 2 en annexe 5.

On s'est d'abord intéressé à la variabilité intersujets du point de vue de la part relative d'évocation de chacun des critères. Cette variabilité est illustrée par le schéma de la figure 5 ci-après.

On constate que la variabilité intersujets diffère selon le critère concerné. Ce sont les critères les plus fréquemment évoqués (Guidage, Signifiante des Codes, Gestion des Erreurs et à un degré moindre Charge de Travail et Consistance) qui sont l'objet de la plus grande variabilité entre sujets. Les catégories les plus faiblement évoquées (Contrôle Utilisateur, Adaptabilité et Compatibilité) sont de ce point de vue beaucoup plus homogènes.

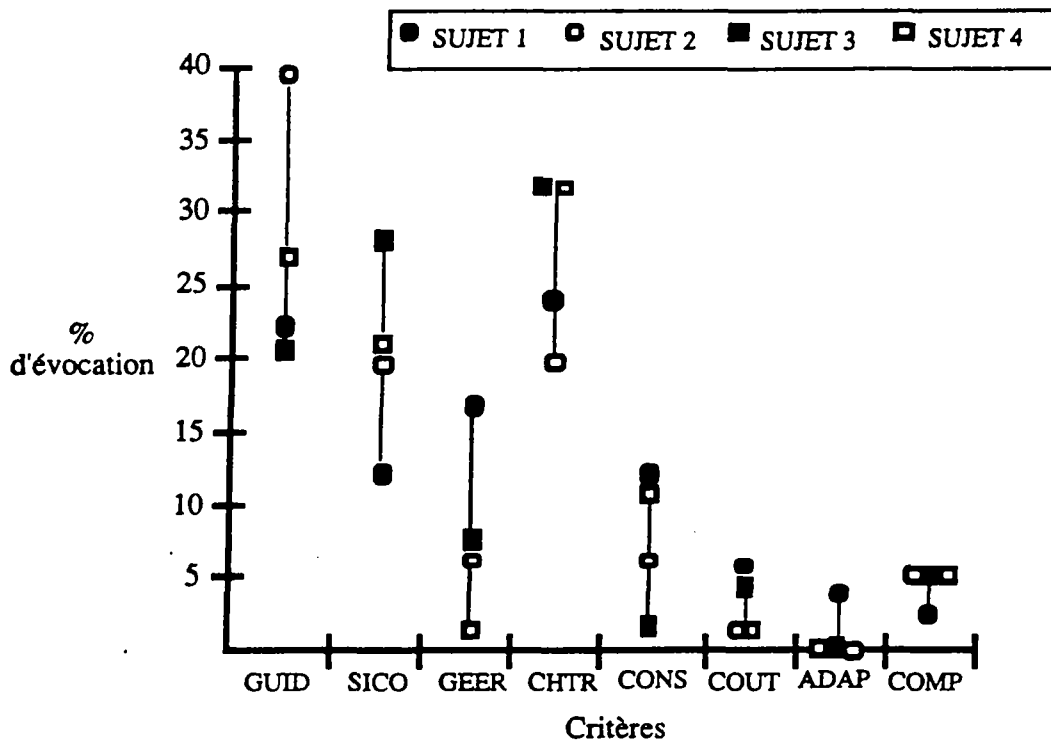


Fig. 5 - Pourcentages relatifs des problèmes évoqués par critère

La variabilité intersujets peut s'expliquer en partie par leurs connaissances préalables du système à évaluer.

En particulier, le sujet 2, peu habitué à manipuler ce type d'interface a été sans doute beaucoup plus sensible aux problèmes de guidage que les autres sujets, beaucoup plus familiers avec les interfaces multifenêtres.

A l'opposé, le sujet 1, ayant une connaissance plus importante de l'application et des notions qui y sont rattachées, a sans doute été beaucoup moins sensible aux problèmes de signifiante des codes que les autres sujets qui découvraient la plupart des termes au cours de l'évaluation.

### 3.4.2.3 Hiérarchie des critères par sujet

L'objectif est ici d'examiner plus précisément la variabilité intersujets du point de vue des priorités d'évocation des différents critères.

Dans cette optique, un calcul de corrélations (coefficient de corrélation de Spearman) a été effectué. Il permet de déterminer dans quelle mesure l'ordre de priorité accordé aux différents critères par chacun des sujets est plus ou moins similaire.

Un coefficient de 1 entre deux sujets signifie que l'ordre de classement des critères est identique pour les deux sujets (un coefficient de 0 signifiant l'absence de relation et un

coefficient de -1 des ordres inverses).

Les résultats obtenus sont résumés dans le tableau 5.

Tableau 5  
Corrélations sur les ordres d'évocation des critères (coefficient de Spearman)

Sujets	Sujet 1	Sujet 2	Sujet 3	Sujet 4
Sujet 1		0,81*	0,73 *	0,70 NS (.053)
Sujet 2			0,81 *	0,89 **
Sujet 3				0,78 *

\* :  $p < .05$     \*\* :  $p < .01$

A partir de ce tableau, on constate une relativement bonne homogénéité entre les sujets du point de vue de l'ordre d'évocation des critères. En effet, toutes les corrélations, mises à part la corrélation entre les sujets 1 et 4, sont significatives et varient entre .73 et .89. Les sujets sont donc relativement proches sur le plan de la hiérarchie des critères.

Néanmoins, ce type de calcul ne prend pas en compte l'importance relative des différents critères évoqués chez un sujet, i.e. la dispersion existant du point de vue de la répartition des problèmes au sein des différents critères. Autrement dit, il ne permet pas d'établir une conclusion quand à la prégnance des hiérarchies.

La prise en compte de cette dimension permet de préciser le résultat précédent. On constate en particulier que la dispersion chez le sujet 1 est relativement plus faible que celle des trois autres sujets :  $\sigma_1 = 8$  ;  $\sigma_2 = 13,44$  ;  $\sigma_3 = 12,55$  ;  $\sigma_4 = 12,52$ .

Les problèmes évoqués par le sujet 1 se répartissent donc de manière plus homogène dans les différentes catégories de critères que ceux évoqués par les sujets 2, 3 et 4. En effet, chez ces derniers, les trois critères Charge de travail, Guidage et Signifiante des codes ont été évoqués beaucoup plus souvent que les autres puisqu'à eux trois ils représentent environ 80% des critères évoqués. Chez le sujet 1 par contre, il y a une prise en compte plus équilibrée de l'ensemble des critères.

Il est difficile d'interpréter ce résultat uniquement sur la base des caractéristiques liées aux sujets. En effet, ce n'est pas seulement le fait que le sujet 1 ait une meilleure connaissance préalable de l'application qui peut expliquer cette relative homogénéité.

Il serait surtout intéressant de voir dans quelle mesure ces différences tiennent à des stratégies différentes d'examen de l'interface.

### 3.4.3 Standards

La catégorie Standards regroupe tous les problèmes caractérisés par les sujets par une non conformité vis-à-vis des standards, en particulier des standards MS-Windows. De façon générale, les Standards représentent une faible proportion de l'ensemble des problèmes évoqués (6%).

Les pourcentages de Standards évoqués par chaque sujet sont présentés dans le tableau 1 en annexe 6. L'histogramme de la figure 6 illustre les résultats obtenus :

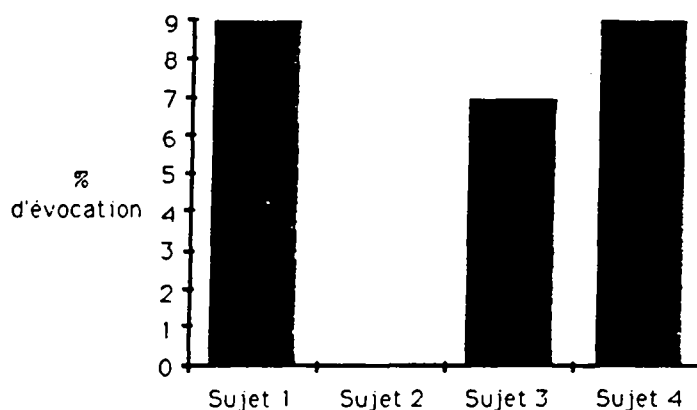


Fig.6 - Pourcentages de Standards évoqués par sujet

La variabilité entre sujets s'explique ici en grande partie par le niveau de connaissance préalable des sujets vis-à-vis des standards MS-Windows. En effet, le sujet 2 découvrait l'environnement MS-Windows lors de l'expérience. Par contre, les autres sujets avaient déjà eu accès à des interfaces de type multifenêtrage (sans pour cela qu'il s'agisse forcément de MS-Windows).

### 3.4.4 Problèmes Fonctionnels

On trouve dans la catégorie Problèmes Fonctionnels, les problèmes qui reposent sur la remise en cause des fonctionnalités mise en œuvre dans l'application. Cette remise en cause peut être de deux types :

- certaines fonctionnalités ont été jugées incomplètes. Par exemple, lorsque l'utilisateur choisit le fichier qu'il va envoyer à un correspondant, il n'a pas la possibilité de pouvoir l'ouvrir pour vérifier son contenu ; ou encore lorsqu'il reçoit un fichier, il ne peut le conserver en mémoire, i.e. le stocker ;
- certaines fonctionnalités ont été regroupées et/ou dissociées à tort : par exemple, l'envoi d'un message de signalisation (ou signalisation usager-usager) est associé à tort avec la fonctionnalité permettant de raccrocher ; ou encore les fonctionnalités "choisir

un fichier" et "transférer un fichier" ont été dissociées à tort (elles sont mises en œuvre dans des contextes différents).

Comme nous l'avons déjà mentionné, on peut supposer que l'évocation de ce type de problèmes est en relation assez directe avec une prise en compte de la tâche utilisateur et que, par conséquent, elle traduit un examen de l'interface à un niveau d'abstraction relativement plus élevé que dans le cas des autres descripteurs (critères et standards).

Les problèmes fonctionnels représentent 11% de l'ensemble des problèmes évoqués.

Les pourcentages de problèmes fonctionnels évoqués par chaque sujet sont présentés dans le tableau 2 en annexe 6.

L'histogramme de la figure 7 illustre les résultats obtenus :

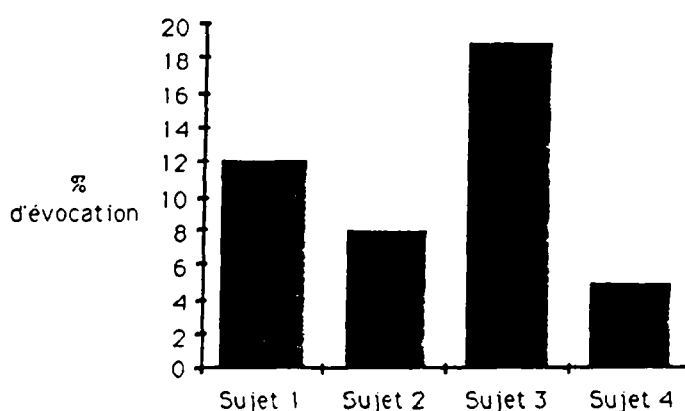


Fig.7 - Pourcentages de problèmes fonctionnels évoqués chez les sujets

On constate que la part de problèmes fonctionnels évoqués est variable selon les sujets. Le sujet 3 a évoqué un grand nombre de problèmes fonctionnels puisque cette catégorie représente près de 20% des problèmes qu'il a diagnostiqués. On trouve ensuite 12% chez le sujet 1, 8% chez le sujet 2 et enfin 5% chez le sujet 4.

Une comparaison de ces résultats avec ceux obtenus lors de l'examen de la variabilité intersujets en fonction des dimensions Affichage vs Séquences de commandes et Existence vs Inexistence ne permet pas de conclure à la présence d'une relation entre ces indicateurs. En effet, on aurait pu penser qu'une proportion importante de problèmes fonctionnels pouvait être liée à une part importante de problèmes appartenant à la catégorie Séquences de commandes et/ou Inexistence.

Par ailleurs, on constate que ce sont les sujets qui ont une expérience plus importante dans le domaine de l'évaluation qui ont évoqué le plus faible pourcentage de problèmes fonctionnels. On pourrait faire l'hypothèse que ceci tient au rôle de l'ergonomie dans la conception de système. En effet, l'évaluation, caractéristique de l'ergonomie

"corrective", intervient souvent dans les phases finales de la conception. A ce stade, les considérations évaluatives ne peuvent généralement que porter sur des aspects de surface, les aspects fonctionnels ne pouvant être remis en question.

### 3.4.5 Interaction entre l'ensemble des descripteurs

L'objectif est ici d'identifier la présence éventuelle d'interactions privilégiées entre l'ensemble des descripteurs (Critères, Standards et Problèmes Fonctionnels) à partir de l'examen des cas où les ergonomes ont attribué spontanément et simultanément deux descripteurs à un problème.

Dans le tableau 6, on a regroupé la totalité des liaisons observées chez les quatre sujets.

Tableau 6  
Nombre de liaisons concomitantes entre descripteurs

Critères	CHTR	COMP	CONS	COUT	ADAP	GEER	GUID	SICO	STAN	PRFO
CHTR		1					7			1
COMP							1	2		
CONS								1	1	
COUT						2				
ADAP										
GEER							3			
GUID								2		
SICO										
STAN										
PRFO										

A partir de ce tableau, on constate que :

- seule une partie des liaisons possibles est apparue, i.e. 10 types de liaisons ont été observés sur les 45 théoriquement possibles ;
- ce sont les critères Guidage (GUID) et Signifiante des Codes (SICO) qui ont été le plus souvent évoqués en interaction avec les autres ;
- l'association "Guidage-Charge de Travail" est la plus fréquente.

L'examen des liaisons observées fait ressortir deux types d'interaction entre critères évoqués simultanément :

- La violation d'un critère entraîne la violation d'un autre.

### Exemples :

- . un mauvais Guidage (prompting) entraîne une augmentation de la Charge de Travail (charge mentale) : le codage couleur d'une option dans un menu laisse penser que cette option est toujours disponible ; or il s'avère en fait que le choix de cette option conduit l'utilisateur à se lancer dans une procédure qui ne peut aboutir, donc à réaliser des actions supplémentaires inutiles ;
- . de façon similaire, l'absence d'indication sur les coordonnées du correspondant lors d'une communication en cours (absence de prompting) oblige l'utilisateur à se souvenir de cette dernière (charge mentale).

Cependant, on a pu observer le phénomène inverse :

- . la violation du critère Charge de travail provoque une détérioration du Guidage : un écran surchargé et/ou la présence d'informations inutiles (charge mentale) peut entraîner des problèmes de guidage (difficulté de savoir dans quel état on est ou de différencier les items actifs / items inactifs) ;
  - . on peut encore citer un autre exemple de liaison de ce type, observé entre les critères Guidage et Gestion des erreurs : la non distinction entre les modes désignation et modes modification d'une liste de données (guidage insuffisant) conduit l'utilisateur à commettre un certain nombre d'erreurs comme la suppression ou le rajout intempestif de données.
- Le deuxième type d'association observé repose sur le fait qu'un problème peut traduire la violation de deux (voir trois) critères à la fois, i.e. sans que l'un soit forcément la conséquence ou la cause de l'autre.

### Exemples :

- . la présence d'une icône inaccessible à l'utilisateur a été présentée à la fois comme un problème de Charge de Travail (présence d'informations inutiles) et un problème de Guidage (icône qui n'est pas différenciée des autres icônes opérationnelles). Deux sujets ont évoqué simultanément ces deux critères, un autre les a évoqués à des moments différents ;
- . la remise en cause du choix d'un terme a été interprétée à la fois comme un problème de Signification des codes (terme peu compréhensible), un problème de Compatibilité (terme ne correspondant pas au vocabulaire habituel des usagers), mais aussi comme un problème de Consistance (sens variant d'un écran à l'autre).



### 3.5 Discussion

L'examen des problèmes évoqués a permis de mettre en évidence une plus ou moins grande homogénéité entre les diagnostics des quatre ergonomes, selon le type d'indicateur utilisé.

En premier lieu, il apparaît que globalement, les performances des sujets sont relativement proches du point de vue de la durée de l'évaluation, du nombre de problèmes évoqués et de l'ordre de priorité d'évocation des différents critères<sup>1</sup>.

De façon générale, ce sont les critères Guidage, Charge de Travail et Signifiante des Codes qui ont été le plus souvent évoqués, les critères Adaptabilité, Contrôle Utilisateur et Compatibilité étant par contre très peu évoqués. Les critères Gestion des Erreurs et Consistance constituent des cas intermédiaires.

La priorité d'évocation des critères est néanmoins plus ou moins prégnante selon les sujets. Chez les sujets 2, 3 et 4, trois critères (Guidage, Charge de Travail et Signifiante des Codes) ont été nettement privilégiés au détriment des autres. Chez le sujet 1, on constate une prise en compte plus équilibrée de l'ensemble des critères.

L'examen des types de problèmes évoqués a permis de constater que tous les sujets ont détecté une part plus importante de problèmes concernant les diverses caractéristiques de l'affichage de l'interface que de problèmes concernant les entrées ou séquences de commandes. De même, les diagnostics d'amélioration (i.e. les problèmes reposant sur une remise en cause d'éléments existants de l'interface) ont été plus nombreux que les diagnostics dénonçant l'absence d'éléments.

Bien que selon les indicateurs précédents les performances des sujets soient globalement proches, le degré de similarité des problèmes évoqués est plutôt faible. En effet, sur l'ensemble des problèmes distincts évoqués, plus de la moitié des problèmes (59%) ont été repérés par un seul des sujets, 21% par deux sujets, 14% par trois sujets, et seulement 6 % par les quatre sujets. Par ailleurs, le degré de similarité des problèmes chez les sujets pris deux à deux a pour maximum 30% et varie selon les paires de sujets.

Bien que surprenants, ces résultats sont proches de ceux obtenus par Hammond et al. (1985) qui comparent les diagnostics d'une même interface effectués par trois paires d'ergonomes expérimentés. En effet, parmi l'ensemble des "points" repérés par les sujets comme posant des difficultés ergonomiques, les pourcentages de "points" se recoupant sont à peu près du même ordre que ceux que nous aurions obtenus avec trois sujets (53% des points sont spécifiques à chacune des paires, i.e. à peu près la moitié des "points" sont communs à au moins l'une des deux paires (dont environ 25%

---

<sup>1</sup> En terme de nombre de fois où un critère est évoqué.

communs à deux paires et 22% aux trois paires)).

Sur l'ensemble des problèmes distincts évoqués, un sujet trouve en moyenne 42% des problèmes de l'interface. Autrement dit, on ne peut se fier à l'évaluation d'un seul sujet pour se faire une idée de l'ensemble des problèmes de l'interface.

Par contre, dès qu'on réunit les diagnostics de deux ou trois sujets, on obtient un échantillon beaucoup plus représentatif des problèmes de l'interface (67% des problèmes en moyenne avec les diagnostics de deux sujets, plus de 80% avec ceux de trois sujets). Il faut donc faire la synthèse des diagnostics de plusieurs évaluateurs (au moins trois) pour avoir une image de l'ensemble des problèmes de cette interface.

Sur ce point, nos conclusions rejoignent celles de Hammond et al. (1985) mais aussi celles de Nielsen et Molich (1990) à l'issue d'une comparaison d'évaluations d'interfaces effectuées cette fois par des concepteurs informaticiens.

Ces résultats constituent un argument important en faveur de la définition d'une méthode favorisant un examen plus exhaustif et systématique des problèmes ergonomiques d'une interface.

Ils nous conduisent également à nous interroger sur les raisons de cette importante variabilité des performances des ergonomes.

Certaines caractéristiques particulières des sujets, liées à leur expérience ou à leurs connaissances préalables concernant l'application peuvent permettre d'expliquer en partie l'importante spécificité des diagnostics effectués.

Par exemple, les résultats concernant la part d'évocation des problèmes caractérisés par les sujets comme non conformes vis-à-vis de Standards (en particulier les Standards MS-Windows) reflètent la plus ou moins grande familiarité des sujets avec l'environnement multifenêtres.

On a constaté également que le pourcentage de problèmes fonctionnels évoqués était relativement faible chez les deux sujets les plus expérimentés en matière d'évaluation, ce qui a été interprété en fonction du rôle de l'ergonomie dite "corrective".

D'un point de vue méthodologique, ceci signifie qu'il faudra à l'avenir porter une attention particulière à la consigne qui est donnée aux sujets. En particulier, il conviendra de préciser à ces derniers s'ils sont autorisés ou non à remettre en cause des décisions de conception de haut niveau (aspects fonctionnels).

En examinant les problèmes détectés spécifiquement par l'un ou l'autre des sujets, on s'aperçoit qu'ils concernent surtout les aspects de l'interface qui sont en rapport avec le domaine d'expérience privilégié des sujets.

Par exemple, le sujet qui a travaillé plus spécifiquement sur les représentations

graphiques a détecté un grand nombre de problèmes se rapportant à la qualité graphique des icônes de cette interface.

De même, un autre sujet, très au courant sur la façon dont les postes (PC) sont interconnectés avec le réseau, a évoqué un nombre de problèmes importants concernant les possibilités de communication offertes au sein de cette application.

Cependant, même si l'expérience ou l'intérêt particulier de chaque sujet joue un rôle important du point de vue de la spécificité des problèmes évoqués, il convient d'examiner dans quelle mesure des stratégies d'examen de l'interface différentes y ont aussi contribué.

Il serait également intéressant de pouvoir déterminer si les problèmes identifiés par les quatre sujets (voir même par trois des sujets) peuvent être considérés comme étant les plus importants, i.e. comme occasionnant les plus grandes difficultés d'utilisation. Pour pouvoir répondre à ce type d'interrogation, il faudrait par exemple demander aux sujets à l'issue de leur diagnostic, de hiérarchiser l'ensemble des problèmes détectés ou de donner une liste des dix problèmes les plus importants en justifiant leurs choix.

Sur le plan méthodologique, la caractérisation des problèmes évoqués selon divers attributs a permis de tirer quelques enseignements quant à la pertinence ou validité de certains des indicateurs utilisés.

Tout d'abord, il semble que les critères d'évaluation, tels qu'ils ont été définis à partir de la littérature (Aschehoug, 1989 ; Scapin, 1990b) permettent une bonne description des problèmes. Dans cette étude, ils ont permis de caractériser globalement 83% de la totalité des problèmes évoqués.

En outre, l'examen de l'affectation simultanée de plusieurs critères à un problème a révélé qu'un petit nombre seulement d'interactions a été exprimé. Celles-ci sont très regroupées et concernent essentiellement les critères Guidage et Charge de travail.

Au-delà de cette étude, il conviendra de valider les définitions de ces critères, notamment en ce qui concerne leur capacité à rendre compte, de façon stable et univoque, de diagnostics d'interfaces variées.

La comparaison des diagnostics des sujets du point de vue du nombre de problèmes évoqués a été parfois problématique en raison de la variabilité du degré de généralité dans l'expression des problèmes. Celle-ci rend aussi difficile le comptage des problèmes lorsqu'on cherche à les distinguer des uns des autres. Il serait donc nécessaire de trouver des moyens qui permettent de mieux contrôler ou prendre en compte cette plus ou moins grande généralité dans l'expression des problèmes.

L'examen des types de problèmes évoqués a permis de mettre en évidence la non

pertinence des dimensions Affichage vs Séquences de Commandes et Existence vs Inexistence pour rendre compte du degré de complexité de l'examen effectué sur l'interface. En effet, on a obtenu des résultats très différents (presque opposés) suivant que l'on s'est intéressé à l'une ou l'autre de ces dimensions. Par ailleurs, aucune relation n'a pu être observée entre une part importante de problèmes appartenant aux catégories Séquences de Commandes et/ou Inexistence et une proportion importante de problèmes fonctionnels.

## 4 STRATEGIES D'EVALUATION

L'analyse précédente concerne la performance d'évaluation du point de vue des problèmes diagnostiqués.

L'aspect de la performance des ergonomes que l'on se propose maintenant d'examiner concerne leur démarche proprement dite, i.e. le séquençement de leur activité, en termes de diagnostics, mais aussi d'examens de l'interface et de prises d'informations.

L'objectif est d'identifier les stratégies des ergonomes, de les caractériser et de les comparer. En particulier, il s'agit d'apporter des éléments de réponse aux questions suivantes : comment les ergonomes organisent l'examen des divers aspects de l'interface, selon quelle(s) dimension(s) et avec quelle(s) priorité(s) ? Peut-on identifier au sein de leur démarche, une approche systématique concernant par exemple, les différents objets de l'interface ou la prise en compte de certains critères ?

On exposera, dans un premier temps, les méthodes d'analyses utilisées pour identifier les stratégies ainsi que les difficultés méthodologiques rencontrées. Les résultats de ces analyses seront ensuite présentés.

### 4.1 Méthodes d'analyses

La présence de stratégies suppose implicitement qu'un certain nombre de règles ou principes directeurs orientent et organisent, tout au moins en partie, la démarche d'évaluation. Du point de vue de l'observable, cela se traduit par un certain nombre de régularités comportementales.

Ces régularités ont été identifiées en examinant la similitude des comportements de chaque ergonome dans différentes situations. Par exemple, les comportements observés dans l'état  $n$  ont été mis en relation avec ceux observés dans l'état  $n+1$ , puis dans l'état  $n+2$ , etc. Ceci nous a permis de voir si certains comportements persistaient ou variaient.

Plusieurs analyses ont été ainsi effectuées en augmentant progressivement la taille des séquences d'évaluation prises en compte pour finalement aboutir à la performance globale.

#### 4.1.1 Caractérisation des comportements élémentaires

Dans un premier temps, chacun des comportements élémentaires observés (examens, actions et contrôles effectués, questions posées et problèmes diagnostiqués) a été caractérisé selon les points de vue suivants :

- Type d'élément concerné : e.g. fonctionnalité, état, objet, transaction, etc.
- Propriété considérée : e.g. format, position, contenu, etc.
- Place de l'élément :
  - . en référence à la structure de l'interface : e.g. niveau dans la hiérarchie de menus ;
  - . en référence à l'écran : e.g. haut/bas, gauche/droite.
- Critère considéré : e.g. guidage, consistance, etc.
- Niveau d'abstraction : e.g. fonctionnel, sémantique, syntaxique ou lexical.
- Degré de généralité : e.g. objet vs classe d'objets.
- Degré de complexité : e.g. transaction élémentaire vs transaction plus complexe, i.e. une combinaison de transactions élémentaires.
- Degré d' "importance" : e.g. fonctionnalités principales (établir une communication) vs fonctionnalités secondaires (modifier les données d'un répertoire).

#### 4.1.2 Analyse des séquences d'évaluation minimales

Dans un second temps, les comportements élémentaires successifs ont été confrontés au niveau des différentes caractéristiques précédemment décrites pour identifier d'éventuelles régularités.

Cette approche a été poursuivie en effectuant un regroupement de comportements élémentaires de plus en plus important, d'une part, pour d'apprécier le caractère plus ou moins systématique des régularités, et d'autre part, pour identifier la présence éventuelle de priorités d'examen, notamment :

- du point de vue des divers objets et de leurs propriétés : e.g. le sujet examine-t-il préférentiellement les boutons de commandes puis les icônes, ou les menus puis les labels ? Le sujet s'intéresse-t-il à la position des objets avant d'examiner leurs formats ? ou inversement ?

- du point de vue des différentes fonctionnalités : e.g. le sujet s'intéresse-t-il aux principales fonctionnalités de l'application avant de considérer des fonctionnalités plus secondaires ?
- du point de vue du degré de généralité des examens effectués, notamment pour les objets : e.g. procède-t-il à un examen global de diverses classes d'objets, comme les fenêtres, avant d'examiner plus précisément leur contenu, ou inversement ? Procède-t-il à un examen de transactions élémentaires avant d'examiner des transactions plus importantes ?
- du point de vue du niveau d'abstraction des examens effectués : e.g. le sujet s'intéresse-t-il aux aspects de l'interface de bas niveaux (lexicaux et syntaxiques) avant de considérer ceux de plus hauts niveaux (sémantiques et conceptuels), ou inversement ?
- du point de vue des critères pris en compte : e.g. le sujet examine-t-il le guidage avant d'examiner la consistance, ou s'intéresse-t-il à la signification des codes avant d'examiner la charge de travail imposée à l'utilisateur ?

#### 4.1.3 Analyse de l'ensemble des séquences d'évaluation

Enfin, on a tenté d'identifier la présence de stratégies en examinant l'ensemble des chemins parcourus par les sujets lors de l'évaluation, i.e. à partir de l'analyse de la succession chronologique des différents états de l'interface explorés.

Ne disposant pas du graphe "théorique" des états de l'interface, on n'a pas été en mesure de raisonner sur l'ensemble des états ou cheminements possibles. En particulier, il n'a pas été possible de déterminer le degré d'exhaustivité de l'examen effectué par les sujets. L'analyse des chemins parcourus a donc été effectuée en fonction des états de l'interface effectivement explorés par les sujets.

Cette analyse s'est avérée extrêmement complexe et a posé un certain nombre de difficultés méthodologiques, en particulier :

- de nombreux problèmes techniques provoquant des réinitialisations du système sont survenus en cours d'évaluation et ont perturbé l'activité spontanée des sujets. Ces problèmes sont apparus dans des situations différentes et ont été plus ou moins importants selon les sujets. En particulier, la situation de double communication (téléphone + données) n'a pu être examinée par un des sujets.
- les interventions de l'expert présent lors de l'expérience n'ont pas été contrôlées. En effet, ce dernier était non seulement présent pour répondre aux questions des sujets

mais aussi pour jouer le rôle du correspondant lors de la simulation des différentes situations de communications, en particulier pour les cas d'appels entrants. Or, il s'est avéré qu'il n'est pas toujours intervenu de la même façon ni au même moment pour tous les sujets. En conséquence, toutes les situations d'interactions n'ont pas été forcément les mêmes pour tous les sujets.

Les chemins complets des sujets n'étant pas strictement comparables, on a recherché des séquences de protocoles non perturbées par les problèmes techniques et qui par ailleurs traduisaient des successions d'états similaires chez tous les sujets.

Cependant, il a été difficile de trouver des séquences communes qui, d'une part correspondaient à des périodes d'évaluation assez proches chez tous les sujets (i.e. permettant de contrôler ce qui avait été fait avant) et d'autre part, étaient suffisamment longues pour être interprétables.

Néanmoins, un examen plus complet des comportements des sujets a été effectué à partir des vingt premiers états explorés, dans la mesure où cette séquence semblait au mieux satisfaire les exigences précédentes. Cette analyse a permis d'illustrer quelques différences interindividuelles importantes sur le plan des stratégies.

Conjointement aux analyses précédentes, on a également examiné l'évolution des diverses caractéristiques des problèmes évoqués au cours de l'évaluation. En particulier, on a tenté de déterminer dans quelle mesure la détection de certains types de problèmes (e.g. problèmes fonctionnels) était liée à une période particulière de l'évaluation.

#### **4.2 Description des stratégies d'évaluation**

Les analyses précédentes ont permis de mettre en évidence qu'une interface pouvait être examinée selon différents points de vue que l'on appellera stratégies d'évaluation. Cinq stratégies ont été identifiées. Certaines sont apparues de façon explicite, notamment à travers les verbalisations des sujets, d'autres sont plus le produit d'inférences et de constructions à partir des comportements observés.

Chacune de ces stratégies va être tout d'abord définie. Ceci nous servira de cadre de référence pour exposer ensuite la façon dont elles ont été mises en œuvre par les ergonomes au sein de la démarche globale d'évaluation.

#### 4.2.1 Stratégie selon des objectifs d'utilisation

Ce type de stratégie consiste à organiser l'examen d'une interface en fonction d'objectifs d'utilisation concrets, i.e. en attribuant des objectifs de tâches à l'utilisateur.

Autrement dit, le diagnostic se fait par le biais de scénarios simulant la réalisation des principales tâches et sous-tâches que l'utilisateur peut être amené à effectuer au travers de l'interface. Cela revient aussi à passer successivement en revue les diverses fonctionnalités ou principaux types de services offerts par l'application et à regarder les problèmes que cela pose.

Au sein d'une stratégie selon des objectifs d'utilisation, plusieurs approches sont possibles suivant les priorités d'examen des différentes fonctionnalités. Par exemple :

- L'examen des fonctionnalités peut se faire par ordre d'importance des tâches utilisateur, e.g. examen des fonctionnalités supportant les tâches principales puis de celles supportant les tâches secondaires.

Dans le cadre de l'interface étudiée, la séquence d'examen pourrait être :

##### 1. Tâches principales :

[Emettre un appel ,  
Transférer un fichier,  
Emettre un appel + Transférer simultanément un fichier ]

##### 2. Tâches secondaires

[Modifier les données des répertoires contenant les diverses caractéristiques des correspondants (numéros d'appel, sous-adresses, etc..)]

- L'examen des fonctionnalités peut aussi se faire par ordre de complexité, e.g. des plus simples aux plus complexes.

La séquence d'examen pourrait être alors :

##### 1. Examen des tâches simples

[Etablir ou Recevoir un type particulier de communication (voix ou données),

Ajouter un numéro d'appel dans la liste des numéros de téléphone]

##### 2. Examen de tâches plus complexes, i.e., une combinaison de tâches simples.

[Etablir une double communication (voix et données),

Remplacer une ligne du répertoire (un numéro de téléphone + une sous-adresse + un message) par une autre]

- Enfin, l'examen des fonctionnalités peut être organisé par une combinaison des deux



séquences précédentes, e.g. examen des fonctionnalités principales, des plus simples aux plus complexes, puis examen des fonctionnalités secondaires, toujours des plus simples aux plus complexes.

#### 4.2.2 Stratégie selon la structure de l'interface

Cette stratégie consiste à organiser l'examen de l'interface en fonction de sa structure.

La structure d'une interface peut être définie comme un arbre théorique représentant l'ensemble des éléments de l'interface en fonction des états qui la composent et des inputs ou actions qui permettent la transition d'un état à un autre.

Selon ce point de vue, l'interface peut être décomposée en "niveaux hiérarchiques" relatifs.

Par exemple, le niveau de base (ou niveau 0) représente l'état ou l'écran d'accueil, i.e. non seulement l'ensemble des informations qu'il contient mais aussi l'ensemble des transactions possibles au sein de cet état. Le niveau 1 représente les nouveaux états résultant de l'activation d'actions (e.g. options de menus, boutons de commandes) accessibles au niveau 0 ; le niveau 2 représente les nouveaux états résultant d'actions accessibles au niveau 1, etc.<sup>1</sup>.

Ce type de représentation est particulièrement bien adapté pour les interfaces comportant une structure hiérarchique de menus.

Le schéma de la figure 8 fournit un exemple d'une telle structure. Pour simplifier, seules les actions provoquant le passage d'un état à un autre de plus bas niveau ont été représentées. En particulier, les actions permettant de revenir aux niveaux précédents ou provoquant des changements à l'intérieur même d'un état, n'y figurent pas.

Au sein d'une stratégie selon la structure, on peut distinguer deux grands types d'approches selon que l'examen se fait plutôt verticalement, i.e. en "profondeur d'abord", ou horizontalement, i.e. en "largeur d'abord".

Une approche en "profondeur d'abord" consiste à examiner chaque branche de l'arbre jusqu'au niveau le plus bas avant d'examiner la suivante.

Par opposition, une approche en "largeur d'abord" consiste à examiner tous les états d'un niveau donné avant d'examiner ceux du niveau suivant.

---

<sup>1</sup> Dans le cadre de l'interface étudiée, on trouvera des exemples d'états appartenant à différents niveaux en annexe 1, notamment : l'écran d'accueil page 2, des états du niveau 1 pages 3 et 4, des états du niveau 2 pages 5 à 10, un état du niveau 3 page 11 de l'annexe 1.

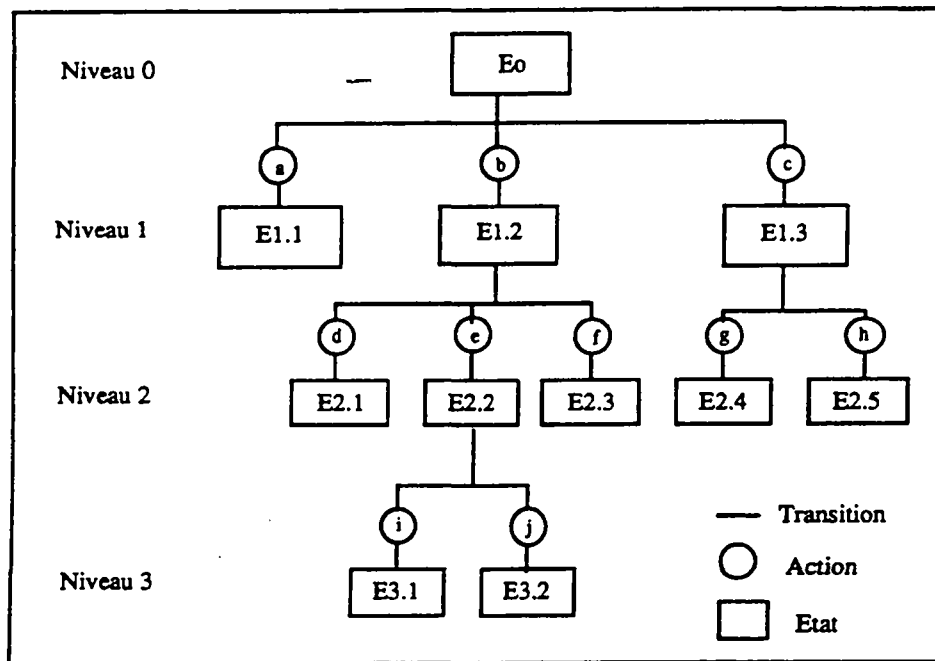


Fig. 8 - Représentation schématique d'une partie de la structure d'une interface

En s'appuyant sur le schéma de la figure 8, une approche en profondeur d'abord se traduirait par la séquence suivante :

[E0 , E1.1, E1.2, E2.1, E2.2, E3.1, E3.2, E2.3, E1.3, E2.4, E2.5].

Une approche en largeur d'abord donnerait :

[E0 , E1.1, E1.2, E1.3, E2.1, E2.2, E2.3, E2.4, E2.5, E3.1, E3.2].

#### 4.2.3 Stratégie selon les niveaux d'abstraction

Cette stratégie consiste à organiser l'examen de l'interface selon différents niveaux d'abstraction, tels qu'ils peuvent être définis à partir des modèles à niveaux classiques (layered models).

Par exemple, Foley et van Dam (1982) distinguent quatre niveaux dans une interface :

- le niveau conceptuel qui comprend les principaux concepts (objets, relations entre objets, opérations sur les objets) utilisés pour définir la tâche de l'utilisateur dans le système. Ce niveau représente les caractéristiques de plus haut niveau de l'interface, i.e. les plus abstraites. ;
- le niveau sémantique qui comprend les fonctionnalités détaillées du système : il détermine l'existence et les effets des fonctions sur leurs objets correspondants ;
- le niveau syntaxique qui définit la séquence des entrées-sorties du système, notamment la grammaire du dialogue (e.g. les règles de séquençement des

commandes) et l'organisation de l'affichage dans le temps et dans l'espace (répartition des différentes aires, positionnement des menus, prompts, messages d'erreurs, etc.) ;

- le niveau lexical qui détermine l'apparence physique des entrées-sorties. Il concerne aussi bien les écrans, le clavier, les dispositifs de désignation que la dénomination des commandes et les modes de codages (fonte d'un texte, largeur et couleur des lignes), etc. Les aspects lexicaux concernent les caractéristiques les plus superficiels ou de plus bas niveaux de l'interface, i.e. les moins abstraites.

Bien que ces niveaux aient été utilisés initialement pour définir un modèle de conception "top-down" de l'interface, ils peuvent également être très utiles pour en organiser le diagnostic.

Au sein d'une stratégie selon les niveaux d'abstraction, on peut distinguer une approche "top-down", qui consiste à examiner les caractéristiques de l'interface des plus abstraites aux moins abstraites, ou inversement, une approche "bottom-up" (Cf. figure 9).

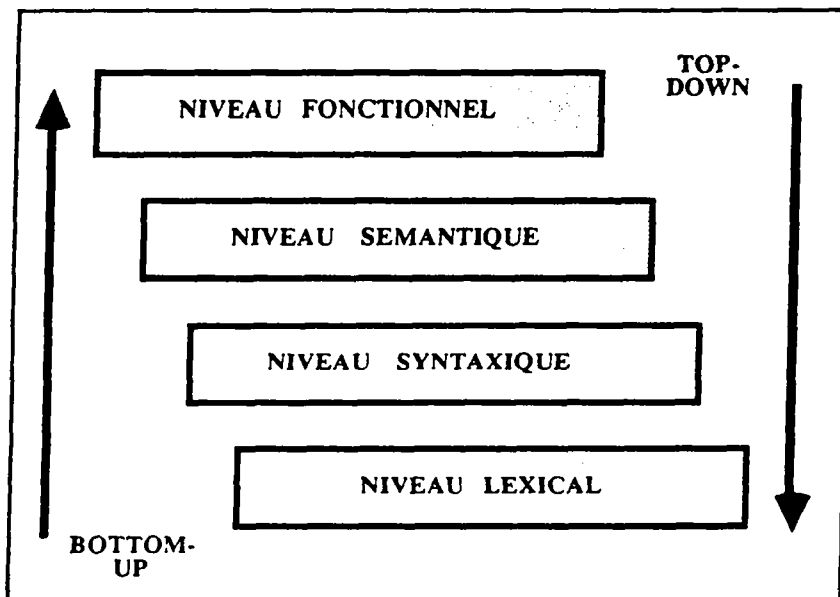


Fig. 9 - Approches Top-down et Bottom-up

Dans le cas d'une approche "Top-down", l'interface sera d'abord examinée du point de vue des différentes fonctionnalités mises en œuvre (capacité fonctionnelle) et de leurs organisations, ainsi que du point de vue de son contenu informationnel. Ensuite, les aspects liés à la syntaxe du langage de commande et à l'organisation des affichages seront considérés avant d'examiner les aspects de surface, tels que le codage externe des options des menus ou des boutons de commandes, l'encadrement des fenêtres, etc.

#### 4.2.4 Stratégie selon les objets de l'interface

Cette stratégie consiste à organiser l'examen de l'interface objet par objet.

L'interface est découpée en objets ou classes d'objets, types d'objets et listes d'attributs (ou propriétés).

Exemple :

(Classe d') objets = [Affichages, Entrées, Dialogues]  
 Types d'objets = { Aires de données, Aires de messages, Aires d'entrées }  
 Attributs = (Structure (composition, position relative), Contenu (style, longueur, dénomination, etc.), Densité, etc.)

Au sein d'une stratégie selon les objets, différentes approches sont envisageables selon l'ordre de priorité d'examen accordé respectivement aux objets, types d'objets et attributs. Par exemple, une séquence d'examen pourrait être :

1. Examen des [Affichages,  
 { Aires de données (Structure, Contenu, etc. ) },  
 { Aires de messages (Structure, Contenu, etc.) },  
 { Aires d'entrées (Structure, Contenu, etc.) }]
2. Examen des [Entrées,  
 { de données (Style, Dénomination, etc.) },  
 { de commandes (Style, Dénomination, etc.) },  
 { type sélection (Style, etc.) }]
3. Examen du [Dialogue,  
 { formulaires (taille des transactions, syntaxe des procédures, etc.) },  
 { menus (nombre d'options, ordre de présentation, codage externe des options, etc.) },  
 { Question/réponse (taille des transactions, syntaxe des procédures, etc.) }]

Une autre séquence pourrait être :

1. [Dialogue,  
 { menus (attributs) }, { questions/réponses () }, { formulaires () }]
2. [Entrées,  
 { type sélection () }, { de commandes () }, { de données () }]
3. [Affichages  
 { Aires d'entrées () }, { Aires de messages () }, { Aires de données () }]

#### 4.2.5 Stratégie selon des critères

Cette stratégie consiste à organiser l'examen de l'interface en fonction d'une liste de critères ergonomiques préétablie<sup>1</sup>. Autrement dit, il s'agit de vérifier successivement si chaque critère ou sous-critère est satisfait.

Par exemple, on peut chercher à vérifier successivement :

le Guidage, i.e.

Prompting,

Feed-back,

Groupement - Distinction entre les items

Clarté

puis,

la Gestion des Erreurs, i.e.

Qualité des messages

Protection contre les erreurs

Correction des erreurs

puis,

la Charge de Travail, i.e.

Brièveté (Concision et Actions minimales)

Charge mentale

puis,

la Consistance, etc.

On imagine facilement, au sein de ce type de stratégie un grand nombre de variantes possibles suivant le nombre de critères et sous-critères considérés et l'ordre d'examen choisi.

#### 4.3 Mise en œuvre des stratégies

Nous allons maintenant tenter de rendre compte de la démarche des ergonomes en nous référant aux stratégies précédemment définies.

Un premier constat concerne la mise en œuvre de l'ensemble de ces stratégies par les sujets. La stratégie selon des objectifs d'utilisation a été adoptée, de façon explicite, par tous les sujets, comme fil conducteur de leur démarche. Par contre, les autres stratégies n'ont pas toujours été mises en œuvre et l'ont été de façon plus ou moins explicite. Ainsi, si trois sujets semblent avoir utilisé la plupart des stratégies au cours de leurs

---

<sup>1</sup> Cf. par exemple la liste de critères proposée au § 3.4.1

diagnostics, cela est beaucoup moins net chez le quatrième, en particulier en ce qui concerne les stratégies selon les niveaux d'abstraction et selon les critères.

Une seconde observation concerne le caractère plus ou moins systématique et exhaustif de la mise en œuvre de ces stratégies. En effet, les régularités comportementales observées ont été plus ou moins fréquentes mais aussi plus ou moins prégnantes, i.e. traduisant parfois plus des tendances comportementales qu'une mise en œuvre rigoureuse de stratégies. Les différentes stratégies ont donc contribué à des degrés et des moments divers à la démarche d'évaluation.

Ceci s'explique en partie par le fait que les sujets étaient non seulement guidés par des stratégies d'évaluation mais aussi par leurs diagnostics précédents et/ou leurs propres difficultés d'utilisation du système.

Par exemple, un sujet s'est donné comme objectif d'examiner la pertinence des labels. Lors de son examen, il remarque à un instant donné le manque de consistance de la position relative des boutons de commandes Quitter/OK. Ceci l'amène à revenir momentanément sur des états précédemment examinés afin de contrôler la position relative de ces mêmes boutons de commandes.

De même, en cours d'évaluation, un sujet se trouve face à une configuration d'écran où plusieurs fenêtres sont superposées. Or il s'avère qu'il ne sait pas comment déplacer les fenêtres les unes par rapport aux autres. Cette difficulté l'amène à examiner toutes les manipulations liées aux mouvements des fenêtres (déplacements, modification de leurs tailles, etc.) et ainsi à dévier momentanément de sa stratégie d'évaluation courante.

Ces observations tendent à confirmer les résultats obtenus lors d'une étude précédente sur l'activité d'évaluation-reconception d'un système de messagerie vocale (Pollier, 1990) Nous avons en effet déjà observé des comportements du même type, caractéristiques d'une démarche opportuniste.

En fait, et c'est sans doute cela le plus important, il s'avère que ces stratégies sont étroitement liées, au sens où elles interagissent de façon emboîtée au sein de la démarche d'évaluation. Autrement dit, ces stratégies n'ont pas été mises en œuvre successivement et indépendamment les unes des autres, mais selon un type d'emboîtement particulier.

La description de la démarche suivie par l'un des sujets va nous permettre d'illustrer la façon dont plusieurs stratégies peuvent être utilisées de façon imbriquée.

La première stratégie mise en œuvre chez le sujet 1 est selon des objectifs d'utilisation avec une revue des différentes fonctionnalités principales, des plus simples aux plus complexes, avec examen au fur et à mesure des fonctionnalités secondaires qui

y sont rattachées.

Au sein de cette première stratégie, s'est inscrite une stratégie selon la structure avec un examen des fonctionnalités qui s'est déroulé plutôt en profondeur d'abord.

La troisième stratégie mise en œuvre est selon des niveaux d'abstraction, avec une approche plutôt de type bottom-up. En effet, ce sujet s'est d'abord intéressé à des caractéristiques lexicales et syntaxiques des affichages et des commandes, puis au contenu sémantique de l'interface, en particulier du point de vue de l'utilité de certains affichages ou commandes, et enfin aux aspects fonctionnels (tous les problèmes fonctionnels ont été détectés en fin de diagnostic). Tout s'est donc passé comme si ce sujet avait effectué une reconstruction d'éléments plus abstraits de l'interface à partir d'éléments moins abstraits.

Au sein de cette stratégie, en particulier au premier niveau d'abstraction, s'est inscrite une quatrième stratégie, selon les objets et leurs propriétés. Le sujet examine d'abord l'affichage puis les inputs ou entrées. Au sein de l'affichage, il considère d'abord les différents éléments (position relative et formats), puis les options de menus et les boutons de commandes (codage externe), etc. Au sein des entrées, il examine d'abord les zones de saisie puis les boutons de commandes, puis les menus, etc. Ce sujet a donc montré un certain nombre de régularités au niveau des éléments examinés, qui, d'un état à l'autre, sont revenues assez régulièrement dans le même ordre.

Enfin, au sein de la stratégie selon les objets, s'est inscrite une stratégie selon des critères d'évaluation. Au niveau des affichages, le sujet vérifie les standards, les critères consistance (localisation et format), guidage (groupement / distinction entre items et prompting), signifiante des codes, et charge de travail (charge mentale). Au niveau des entrées et séquences des commandes, le sujet vérifie les standards, les critères guidage (feed-back), gestion des erreurs, adaptabilité et charge de travail (actions minimales).

La démarche adoptée par le sujet 1 est résumée dans le tableau 7.

L'examen des protocoles des autres sujets tend à confirmer l'existence de ce type d'emboîtement entre stratégies.

Cependant, des différences notables sont apparues du point de vue des approches ou priorités d'examen adoptées au sein de ces stratégies. Par exemple, si les démarches des sujets 1 et 4 sont relativement proches du point de vue des stratégies mises en œuvre, leurs approches ont été diamétralement opposées.

Tableau 7

Mise en œuvre des stratégies d'évaluation : démarche du sujet 1

Stratégie selon les objectifs utilisateurs	Stratégie selon la structure	Stratégie selon les niveaux d'abstraction	Stratégie selon les objets (et leurs attributs)	Stratégie selon des critères
Séquence d'examen des Scénarios ou Fonctionnalités	Approche utilisée	Approche utilisée	Séquence d'examen des objets	Critères pris en compte
<ul style="list-style-type: none"> <li>- Etablir une communication téléphonique</li> <li>. modifier les répertoires</li> <li>. rappeler un correspondant</li> <li>. émettre un appel à partir d'une application</li> <li>- Etablir un transfert de fichier</li> <li>- Etablir une double communication (téléphone et transfert de fichiers)</li> </ul>	Profondeur d'abord	Bottom-up	<u>Affichages</u>	<ul style="list-style-type: none"> <li>- Standards ;</li> <li>- Guidage (Groupement/Distinction entre items)</li> <li>- Consistance (localisation et format)</li> </ul>
		Lexical	<ul style="list-style-type: none"> <li>- Affichages de données / aires d'entrée / aires de messages (position relative, format)</li> <li>- Options de menus, boutons de commandes (codage externe et signification)</li> </ul>	<ul style="list-style-type: none"> <li>- Standards</li> <li>- Guidage (Prompting)</li> <li>- Signifiante des codes</li> </ul>
		et		
			<u>Entrées</u>	<ul style="list-style-type: none"> <li>- Guidage (Feed-back)</li> <li>- Gestion des erreurs</li> <li>- Standards</li> </ul>
			<ul style="list-style-type: none"> <li>- de données : zone de saisie (contexte d'utilisation correct puis incorrect =&gt; filtrage des caractères incohérents ou dépassement de la zone de saisie ; standards d'édition)</li> </ul>	
		Syntaxique	<ul style="list-style-type: none"> <li>- sélection: boutons de commandes et options de menus (contexte d'utilisation correct puis incorrect ; équivalents claviers et raccourcis autorisés)</li> </ul>	<ul style="list-style-type: none"> <li>- Guidage (Feed-back)</li> <li>- Gestion des erreurs</li> <li>- Contrôle utilisateur (actions explicites)</li> <li>- Adaptabilité</li> <li>- Standards</li> </ul>
		Sémantique	<u>Affichages</u>	<ul style="list-style-type: none"> <li>- Charge de travail (charge mentale)</li> <li>- Guidage (prompting)</li> <li>- Signifiante des codes</li> </ul>
		et	<ul style="list-style-type: none"> <li>- Fenêtres (contenu informationnel)</li> <li>. informations nécessaires et suffisantes ?</li> <li>. au bon endroit ?</li> <li>. titre des fenêtres véhicule t-il ce qu'elles représentent ?</li> </ul>	
			<u>Séquences de commandes et dialogue</u>	<ul style="list-style-type: none"> <li>- Charge de travail (Actions minimales)</li> <li>- Compatibilité</li> </ul>
		Conceptuel	<u>Fonctionnalités</u>	
			<ul style="list-style-type: none"> <li>- sont-elles complètes</li> <li>- sont-elles correctement organisées ?</li> </ul>	



Ainsi, au sein d'une stratégie selon des objectifs d'utilisation, le sujet 4 a d'abord examiné globalement toutes les fonctionnalités principales, i.e. les différents types de communications possibles, avant de considérer des fonctionnalités plus secondaires, e.g. modifier les données des répertoires.

L'examen des différentes fonctionnalités s'est déroulé plutôt en largeur d'abord et d'autre part selon une approche de type top-down. En effet, ce sujet s'est intéressé en priorité aux aspects sémantiques et syntaxiques de l'interface. Les aspects de plus bas niveaux (lexicaux), tels que les couleurs, formats ou qualités esthétiques des affichages ont été considérés au cours de la dernière période de l'évaluation.

Sur le plan des objets et de leurs propriétés, ce sujet a examiné en priorité le contenu informationnel des affichages ainsi que le nombre et la longueur des transactions nécessaires pour accomplir les principales tâches utilisateurs. Les affichages et entrées individuels ont été ensuite considérés.

Les critères évoqués prioritairement concernent surtout la charge de travail (charge mentale et actions minimales), la compatibilité, le guidage (prompting et feed-back) et la signifiante des codes (en particulier pour le titre des fenêtres). Les critères consistance (localisation et format), guidage (distinction-groupement entre items et clarté) ont été pris en compte plutôt en fin de diagnostic.

En inversant le tableau 7 (lecture de bas en haut), on obtient une image assez proche de la démarche adoptée par le sujet 4.

La démarche du sujet 3 est globalement assez proche de celle du sujet 4. L'approche qu'il a adoptée au sein de la stratégie selon les niveaux d'abstraction est toutefois singulière. En effet, si ce sujet a examiné en priorité, comme le sujet 4, des aspects sémantiques, puis des aspects syntaxiques et lexicaux, il a cependant terminé son diagnostic en considérant des aspects de hauts niveaux de l'interface, concernant notamment la complétude des fonctionnalités mises en œuvre dans l'application. Bien que majoritairement top-down, ce type d'approche peut être qualifié de "mixte" dans la mesure où la dernière phase de l'évaluation est ascendante du point de vue des niveaux d'abstraction. Un phénomène un peu analogue a été observé au sein de la stratégie selon la structure. En effet, on a constaté une tendance chez ce sujet à examiner chaque fonctionnalité alternativement en largeur d'abord puis en profondeur.

La démarche du sujet 2 se rapproche de celle du sujet 1. En effet, le sujet 2 a débuté son examen des fonctionnalités plutôt en profondeur d'abord et selon une approche bottom-up. Cependant, en dehors de la stratégie selon les objectifs d'utilisation à laquelle ce sujet semble toujours s'être référé, les tendances comportementales observées au cours de la première période se sont estompées pour laisser place à une démarche beaucoup

moins structurée, en majeure partie guidée par les données, i.e. par les diagnostics précédents et/ou difficultés d'utilisation.

Ceci n'est pas étranger au fait que le sujet 2, contrairement aux autres, découvrait l'environnement MS-Windows et les manipulations de la souris. Les effets d'apprentissage ont interféré avec l'activité d'évaluation, multipliant les difficultés d'utilisation, les erreurs et les réinitialisations système. Ce type de situation est sans aucun doute peu favorable à la mise en œuvre et au maintien de stratégies.

#### 4.4 Discussion

A l'issue de l'examen des stratégies, on constate que la notion de démarche d'évaluation est particulièrement complexe et en tout cas, qu'elle ne peut se résumer à la mise en œuvre d'une seule stratégie commune et immuable à tous les sujets. Cinq stratégies très imbriquées, correspondant à des examens de l'interface selon des points de vue différents, ont contribué à des degrés et des moments divers à la démarche d'évaluation.

Ces stratégies ne sont pas juxtaposées mais elles s'articulent selon un type d'emboîtement qui est apparu relativement stable.

La stratégie selon des objectifs d'utilisation permet un premier découpage global de l'examen de l'interface en différents scénarios ou fonctionnalités. Ces derniers correspondent aux principales tâches ou sous-tâches que l'utilisateur peut effectuer à travers l'application.

La stratégie selon la structure détermine ensuite selon quelle séquence d'examen ces différents scénarios ou fonctionnalités seront passés en revue, e.g. en profondeur ou en largeur d'abord.

Au sein de ces examens, la mise en œuvre d'une stratégie selon des niveaux d'abstraction détermine les aspects à considérer (e.g. sémantiques, syntaxiques, etc.) et selon quelle priorité (e.g. approches top-down ou bottom-up).

La stratégie selon les objets, en interaction avec la stratégie selon les critères, organise ensuite l'examen de ces aspects en termes d'objets, de propriétés et de critères d'évaluation.

Compte tenu de ce mode d'articulation entre stratégies, les premières stratégies jouent un rôle déterminant puisqu'elles limitent en grande partie les éléments à considérer au sein des stratégies sous-jacentes. Autrement dit, les dernières stratégies ne s'appliquent plus à l'ensemble de l'interface mais seulement à certains aspects ou éléments, résultant de la mise en œuvre des stratégies précédentes.

Dans cette étude par exemple, la stratégie selon des niveaux d'abstraction est mise en œuvre dans le cadre de l'examen d'une ou plusieurs fonctionnalités particulières.

La stratégie selon les objets ne s'applique qu'aux objets et attributs concernés dans le cadre de l'examen d'une fonctionnalité particulière et à un niveau d'abstraction donné.

En réalité, cet enchaînement de stratégies n'est pas aussi rigoureux et systématique. En effet, les comportements des sujets ne sont pas seulement orientés par des stratégies prédéfinies mais aussi, à des degrés divers, par les diagnostics précédents et/ou leurs propres difficultés d'utilisation du système. Ce type d'observation, témoin d'aspects opportunistes dans la démarche d'évaluation, tend à confirmer les résultats obtenus récemment dans le domaine de la conception (Pollier, op. cit.).

Même si les sujets ont utilisé de façon plus ou moins systématique et exhaustive les stratégies, il est important de noter que la stratégie selon des objectifs d'utilisation a été adoptée explicitement par tous les sujets comme fil conducteur de leur démarche. Ceci rejoint également les observations réalisées lors d'une précédente étude (Pollier, op. cit.). En effet, l'analyse de plans de séance pour l'évaluation du dialogue d'une messagerie vocale avait déjà permis de constater que l'examen s'organisait d'abord autour d'un certain nombre de scénarios et fonctionnalités à tester. Les dimensions objets, critères et niveaux d'abstraction contribuaient également de façon plus ou moins explicite à l'organisation de l'examen du dialogue.

Des variations interindividuelles importantes sont apparues du point de vue des approches adoptées au sein des stratégies. Deux profils assez opposés ont été observés. Le sujet 1 et à un moindre degré le sujet 2 ont adopté une approche en profondeur d'abord (stratégie selon la structure) en association avec une approche bottom-up (stratégie selon les niveaux d'abstraction). Le sujet 4 et à un moindre degré le sujet 3 ont plutôt opté pour une approche en largeur d'abord associée à une approche top-down. Au sein d'un même profil, on a aussi pu relever certaines variations notamment la possibilité de mixer les approches bottom-up et top-down (sujet 3).

Ces variations interindividuelles ne peuvent être expliquées par le biais du facteur expérience. En effet, les sujets ayant des approches similaires ont un niveau d'expérience très variable en matière d'évaluation ergonomique.

Par ailleurs, il n'existe pas de lien direct entre les proximités des sujets établies à partir des stratégies et celles établies globalement par nombre de problèmes communs. En effet, les deux sujets ayant la plus forte proportion de problèmes en communs (31%), i.e. les sujets 1 et 3, ont adopté des approches différentes. Les sujets 3 et 4 qui partagent aussi un nombre relativement important de problèmes (27%) sont par contre plus proches sur le plan stratégique.

Nous avons vu à l'issue de l'examen des problèmes évoqués que les diagnostics spécifiques aux sujets étaient liés en grande partie à leur domaine d'expérience privilégié.

Le même type de conclusion s'impose en matière de stratégies dans la mesure où les aspects examinés avec le plus d'attention sont le plus souvent en relation avec le domaine de connaissances privilégié des sujets. L'importante spécificité des problèmes semble donc plus liée aux centres d'intérêts particuliers des sujets qu'à des différences en termes de stratégies ou d'approches.

Cependant, certaines configurations de l'interface ont aussi fait l'objet d'un examen particulièrement minutieux parce qu'elles représentaient des situations d'interactions auxquelles les utilisateurs risquaient d'être fréquemment exposés.

Le sujet 4, notamment, a examiné avec une attention particulière la fenêtre contenant la liste des messages servant à la signalisation usager-usager (ce sont les messages que l'on envoie aux correspondants en début et en fin de communication). Pour ce sujet, les problèmes liés à la modification de la liste des messages ont été jugés très importants du fait qu'ils concernent une tâche que l'utilisateur risque d'effectuer très fréquemment.

Sur le plan pratique, ce premier examen des stratégies suggère qu'une assistance à l'évaluation ergonomique d'interfaces, qu'elle soit automatisée ou non, doit être vue comme un environnement très flexible autorisant l'examen d'une interface selon plusieurs points de vue : selon des objectifs d'utilisation, selon des objets, des niveaux d'abstraction (avec possibilités ascendantes et descendantes), selon la structure (vues en largeur ou en profondeur d'abord), mais aussi selon des critères.

Pour aboutir à une définition plus précise d'une méthode ou démarche d'évaluation, un certain nombre de problèmes doivent être préalablement considérés, notamment :

- la nécessité de disposer de modèles et de concepts suffisamment bien définis pour étayer une éventuelle méthode d'investigation de l'interface. En particulier, il est important de progresser dans le domaine de la modélisation des niveaux d'abstraction, de la structure et des objets de l'interface, mais aussi dans le domaine de la définition-validation de critères et d'objectifs d'utilisation pour l'évaluation.
- un second point concerne la validité des stratégies définies dans cette étude, du point de vue de leur généralité et de leur applicabilité.

Concernant le premier point, il apparaît déjà que certaines stratégies semblent être des points de référence relativement stables pour organiser l'examen d'une interface. En effet, les dimensions objectifs d'utilisation, objets, critères et de façon moins explicite, niveaux d'abstraction ont par ailleurs servis à élaborer des plans d'évaluation d'un

système de messagerie vocale (Pollier, op. cit.).

La stratégie selon la structure semble par contre plus particulièrement adaptée aux interfaces comprenant une structure de menus hiérarchisés, i.e. dans laquelle l'utilisateur doit naviguer à travers plusieurs niveaux.

Ces premiers constats doivent de toute façon être complétés par d'autres expériences de ce type, avec d'autres experts et sur des interfaces de types variés.

Sur le plan de l'applicabilité, il s'agit de pouvoir tester la possibilité de faciliter et d'améliorer le diagnostic d'une interface en organisant son examen selon les divers points de vue suggérés par les stratégies, à la fois auprès d'ergonomes mais aussi d'informaticiens peu expérimentés dans ce domaine.

## 5 CONCLUSION

Cette étude a permis un premier examen de l'activité de quatre ergonomes expérimentés en situation d'évaluation individuelle d'une interface existante.

Deux aspects de la performance d'évaluation ont été examinés : d'une part, les résultats des diagnostics, i.e. les problèmes évoqués, et d'autre part, la façon dont l'examen des divers aspects de l'interface a été organisé, i.e. les stratégies.

L'examen des problèmes évoqués a permis de mettre en évidence que la performance des sujets varie selon les aspects envisagés.

Sur le plan de la durée de l'évaluation, du nombre de problèmes évoqués et des priorités d'évocation des critères ergonomiques, les sujets sont relativement proches.

Cependant, le degré de similarité des problèmes est plutôt faible puisque sur l'ensemble des problèmes distincts évoqués, moins de 10% sont communs aux quatre sujets (14% à trois sujets, 21% à deux sujets et 59% de problèmes spécifiques). La proportion de problèmes communs chez les sujets pris deux à deux ne dépasse pas un tiers (en moyenne 25%).

Si l'on associe à cette importante spécificité des problèmes le fait qu'un sujet trouve en moyenne moins de la moitié (42%) des problèmes de cette interface, on constate qu'il faut faire la synthèse des diagnostics d'au moins trois évaluateurs pour avoir une image de l'ensemble des problèmes de cette interface. En rassemblant les diagnostics de deux sujets, on obtient en moyenne deux tiers des problèmes.

L'examen des stratégies a permis de mettre en évidence que la notion la démarche

d'évaluation ergonomique est particulièrement complexe et, en tout cas, qu'elle ne peut se résumer à la mise en œuvre d'une seule stratégie commune à tous les sujets.

Cinq stratégies très imbriquées ont contribué à des degrés et à des moments divers à l'organisation de l'examen de l'interface : stratégies selon des objectifs d'utilisation, selon la structure de l'interface, selon des niveaux d'abstraction, selon les objets de l'interface, et enfin selon des critères.

Ces stratégies n'ont pas été mises en œuvre successivement et indépendamment les unes des autres, mais selon un type d'emboîtement qui est apparu relativement stable.

La stratégie selon des objectifs d'utilisation permet un premier découpage global de l'examen de l'interface en différents scénarios ou fonctionnalités correspondant aux principales tâches ou sous-tâches de l'utilisateur. La stratégie selon la structure détermine ensuite selon quelle séquence d'examen ces différents scénarios ou fonctionnalités seront passés en revue, e.g. en profondeur ou en largeur d'abord. Au sein de ces examens, la mise en œuvre d'une stratégie selon des niveaux d'abstraction détermine les aspects à considérer (e.g. sémantiques, syntaxiques, etc.) et selon quelle priorité (e.g. approches top-down ou bottom-up). La stratégie selon les objets en interaction avec la stratégie selon les critères, organise ensuite l'examen de ces aspects en termes d'objets, de propriétés et de critères d'évaluation.

Des variations interindividuelles importantes sont apparues sur le plan des approches ou priorités d'examen adoptées au sein de ces stratégies.

Deux sujets ont choisi d'examiner les fonctionnalités plutôt en profondeur d'abord (stratégie selon la structure) et selon des niveaux d'abstraction de plus en plus élevés (approche bottom-up). Les deux autres ont adopté une démarche inverse : ils ont réalisé un examen plutôt en largeur d'abord et selon des niveaux d'abstraction de moins en moins élevés (approche top-down).

Ces variations ne peuvent s'expliquer par l'expérience des sujets en ergonomie. Par ailleurs, elles ne peuvent rendre compte de la plus ou moins grande proximité des sujets du point de vue des problèmes évoqués conjointement.

A l'issue de ces résultats, on est amené à s'interroger sur les raisons de cette importante variabilité des performances des ergonomes, en particulier du point de vue des problèmes évoqués.

La situation d'expérience a certainement joué un rôle important. En effet, on peut penser qu'on aurait obtenu une plus grande homogénéité entre les diagnostics des ergonomes en les mettant dans des conditions plus proches de leurs conditions habituelles de travail, e.g. en leur donnant la possibilité de réexaminer l'interface après un certain laps de temps (e.g. une journée).

Cependant, il n'en reste pas moins que les caractéristiques particulières des sujets, liées à leurs domaines d'expérience et à leurs connaissances préalables concernant le produit à évaluer ont joué un rôle déterminant dans la diversité des diagnostics. Même si les évaluations d'experts sont efficaces, il existe tout de même un biais de focalisation important.

Ce phénomène n'est ni nouveau, ni spécifique aux ergonomes. En effet, des résultats du même ordre ont été obtenus précédemment (Hammond et al., op. cit.), non seulement en confrontant les diagnostics d'une même interface fournis par différents experts en ergonomie mais aussi en confrontant les évaluations issues d'observations d'utilisateurs novices. Une importante spécificité des problèmes a été également mise en évidence lors d'évaluations réalisées par des informaticiens (Nielsen et Molich, op. cit.). De plus, ces auteurs constatent également que les performances d'évaluation d'une même personne varient d'une interface à l'autre.

Par ailleurs, l'examen des stratégies montre qu'en dépit du fait que les ergonomes ont mis en œuvre, à des degrés d'élaboration divers, un certain nombre de stratégies, il apparaît clairement que leur approche n'est ni systématique ni exhaustive et qu'ils ont une démarche de type opportuniste.

Toutes ces observations montrent que le processus d'évaluation est complexe et confirment la nécessité actuelle de définir une méthode qui permette une identification plus systématique et plus exhaustive des problèmes ergonomiques d'une interface, et moins dépendante de l'expérience et de l'intérêt particulier des évaluateurs.

Le problème est alors de savoir quelles doivent être les principales caractéristiques et modalités d'une méthode (ou d'un environnement d'assistance) pour l'évaluation ergonomique d'interfaces.

De façon générale, les diverses stratégies identifiées dans cette étude conduisent déjà à envisager l'aide à l'évaluation non pas sous la forme d'un seul algorithme ou d'un seul système expert mais sous la forme d'un environnement flexible autorisant un examen de l'interface selon divers points de vue et avec différentes priorités.

Cependant, pour aller au-delà de cette perspective générale, on peut d'ores et déjà s'interroger sur trois points :

Le premier concerne la nécessité de faire appel à l'ensemble des points de vue identifiées dans cette étude.

Nous avons constaté que la plupart des stratégies semble déjà des points de référence relativement stables pour organiser l'examen d'une interface, même si la stratégie selon la structure peut apparaître comme étant plus adaptée aux interfaces utilisant un

dialogue basé sur des menus hiérarchisés.

Nous pensons par ailleurs que ces divers points de vue sont complémentaires et nécessaires. En effet, pour spécifier une démarche permettant un examen exhaustif et systématique de l'interface, i.e. permettant de définir à la fois quels aspects ou éléments de l'interface doivent être examinés, dans quel ordre et selon quels critères, il faudrait employer de façon emboîtée les diverses stratégies mises en évidence dans cette étude.

Un second point concerne la manière de coordonner ces points de vue.

On peut se baser sur le type d'emboîtement observé dans cette étude ou en envisager d'autres. En particulier, on peut déjà a priori éliminer les emboîtements et approches qui semblent théoriquement difficile à envisager. Par exemple, une stratégie selon des objectifs d'utilisation peut difficilement s'insérer au sein d'une stratégie selon les objets. Par ailleurs, on peut se demander si un examen en profondeur d'abord est aussi compatible avec une approche bottom-up qu'avec une approche top-down.

Le troisième point, corollaire du précédent, est de savoir quel(s) emboîtement(s) et approche(s) il est souhaitable de favoriser. Autrement dit, parmi les possibilités envisageables, y-en-a-t-il une qui permette un diagnostic plus efficace, i.e. à la fois plus exhaustif (en identifiant le maximum de problèmes) et plus économique (en optimisant le rapport entre l'exploration et les problèmes détectés).

Cette étude ne nous permet pas de nous prononcer à ce sujet, d'autant plus que les sujets ayant la plus grande proportion de problèmes en commun ont adopté des approches différentes.

On a constaté par ailleurs que les divergences d'approches observées ne peuvent s'expliquer par le biais de l'expérience en ergonomie puisque les sujets ayant des approches plutôt similaires ont un niveau d'expérience très variable en matière d'évaluation. Autrement dit, le choix d'une approche particulière (e.g. approche top-down et approche en largeur d'abord) semble indépendant de l'expérience des sujets en ergonomie.

Pour apporter des éléments de réponses plus précis à ces trois points, il semble nécessaire de poursuivre cette étude en examinant d'autres experts, dans des environnements d'interfaces variés, posant d'autres types de problèmes ergonomiques et ceci dans une double perspective.

- d'une part, pour apprécier le degré de généralité des stratégies et de leur emboîtement identifiés dans cette étude. En particulier détecte-t-on d'autres formes d'organisation ?
- et d'autre part, sur un plan plus pratique, pour tester la possibilité d'améliorer le



diagnostic d'une interface en organisant son examen à l'aide des points de vue mis en évidence dans cette étude.

Au delà de cette étude, plusieurs aspects doivent être approfondis ou restent à explorer, notamment :

- la validation des critères et leur apprentissage. Il s'agit d'examiner la stabilité d'affectation de diagnostics ergonomiques vis-à-vis des critères utilisés dans cette étude. Dans cette perspective, un premier travail de validation des critères et de leurs définitions a commencé (Bastien, 1990). Par ailleurs, il conviendra d'examiner si ces critères peuvent être fournis à des concepteurs comme des moyens d'évaluation du caractère ergonomique d'une interface ;
- la distinction entre évaluation et exploration. Il serait intéressant de pouvoir expliciter cette distinction afin d'envisager une méthode générale d'examen de l'interface, indépendante du type d'application envisagé. A cet effet, on peut penser à des expériences de description d'interfaces dans une situation d'échange d'informations entre spécialistes en ergonomie.

Les difficultés méthodologiques rencontrées dans cette étude nous permettent de formuler un certain nombre d'indications utiles pour nos futurs travaux.

Concernant l'examen des diagnostics, il serait souhaitable de pouvoir :

- déterminer préalablement le nombre et les caractéristiques des problèmes effectivement présents au sein de l'interface, e.g. à l'aide d'experts ne participant pas à l'expérience mais aussi de façon empirique, i.e. avec des utilisateurs potentiels ;
- contrôler la généralité mais aussi l'importance des problèmes évoqués. A cet effet, on pourrait par exemple demander aux sujets à l'issue de leurs diagnostics, de fournir une liste des problèmes qu'ils considèrent comme les plus importants, en justifiant leurs choix. Une autre façon de procéder, plus proche d'une situation habituelle de travail, serait de demander aux ergonomes de fournir un rapport écrit mettant en évidence les aspects importants ou résumant les principaux défauts de l'interface. Ceci aurait en plus l'avantage d'aborder le problème du diagnostic global de la qualité d'une interface à partir de l'identification d'un ensemble de problèmes particuliers ;
- contrôler les objectifs d'évaluation des sujets. En particulier, il faudrait mettre les sujets dans une situation qui les incite à effectuer un examen plus large, i.e. où ils s'autorisent à remettre en cause tous les aspects de l'interface, notamment ceux de niveau conceptuel.

Concernant l'accès et l'examen des stratégies, il serait utile d'affiner la grille d'analyse des comportements qui a été déterminée empiriquement dans cette étude. En particulier, il serait souhaitable d'introduire des descripteurs quantifiables, permettant d'aller au-delà d'une simple description qualitative des stratégies. Par ailleurs, un meilleur contrôle de l'exhaustivité de l'examen effectué pourrait être réalisé en définissant au préalable un graphe des cheminements possibles au sein de l'interface.

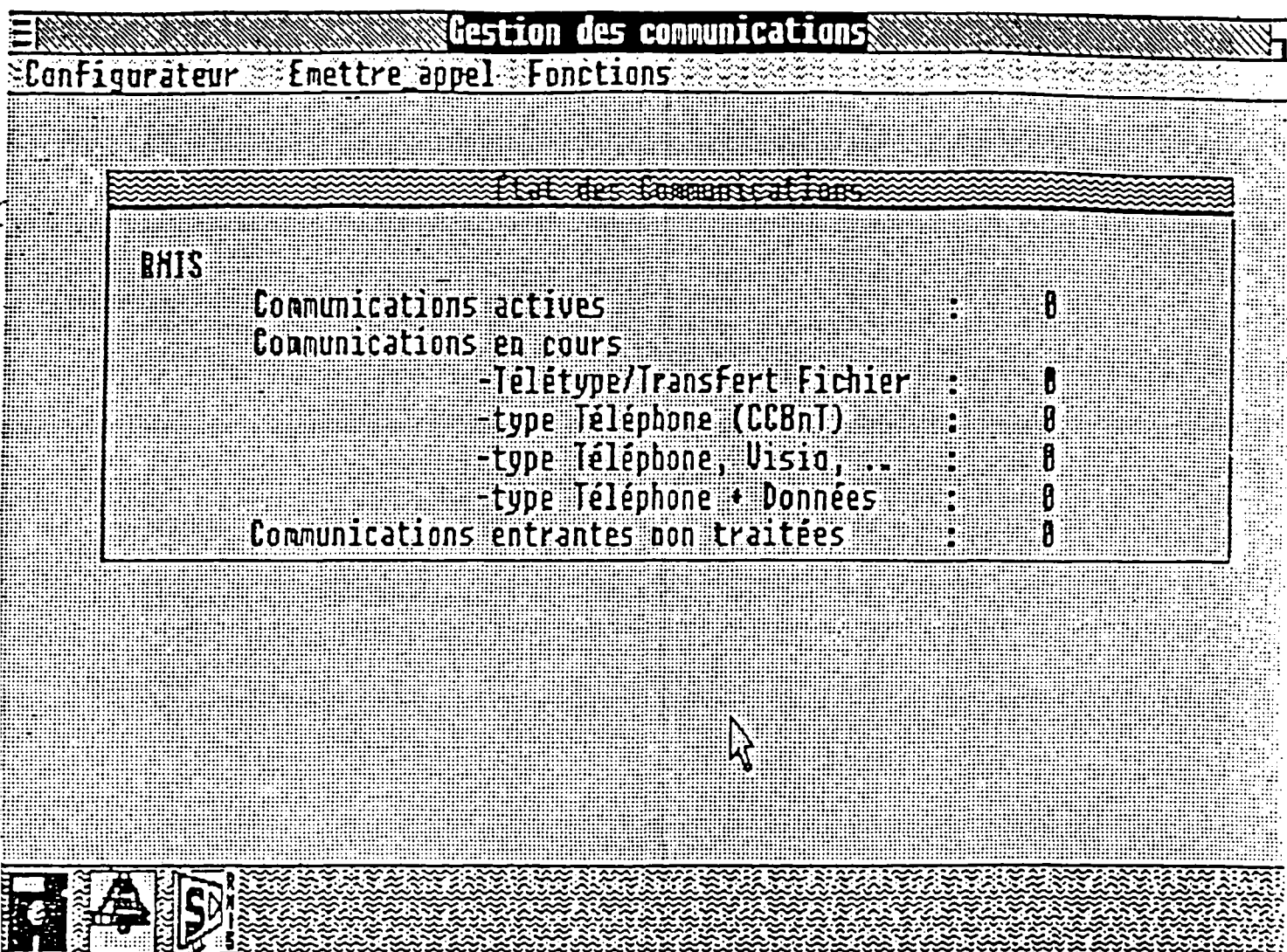
## REFERENCES

- Aschehoug, F. (1989) - *Construction d'une base de recommandations pour le diagnostic ergonomique des interfaces multifenêtres*. Rapport Convention CNET, Novembre 1989.
- Bastien, C. (1990) - *Validation de critères ergonomiques pour l'évaluation d'interfaces utilisateurs*. Rapport de Recherche INRIA (à paraître).
- Card, S., Moran, T.P. et Newell, A. (1983) - *The psychology of human-computer interaction.*, Hillsdale, N.J. : Lawrence Erlbaum Associates.
- Carroll, J.M. et Rosson, M.B. (1985) - Usability Specifications as a Tool in Iterative Development. In *Advances in Human-Computer Interaction*, H. R. Harston (Ed.), Vol 1.: Norwood, NJ : Ablex Publishing Corporation, 1-28.
- Foley, J.D. et van Dam, A. (1982) - *Fundamentals of Interactive Computer Graphics*. Reading, MA: Addison-Wesley Publishing Company.
- Gould, J.D. et Lewis, C. (1985) - Designing for Usability : Key Principles and What Designers Think. *Communication of ACM*, 28, 300-311.
- Hammond, N., Hinton, G., Barnard, P., MacLean, A., Long, J. et Whitefield, A. (1985) - Evaluating the interface document processor : a comparison of expert judgement and user observation. In *Human-Computer Interaction, INTERACT'84*, B. Shackel (Ed.), Amsterdam, North-Holland : Elsevier Science Publishers, IFIP, 725-729.
- Howard, S. et Murray, D.M. (1987) - A taxonomy of evaluation techniques for HCI. In *Human-Computer Interaction, INTERACT'87*, H. J. Bullinger et B. Shackel (Eds.), Amsterdam, North-Holland : Elsevier Science Publishers, IFIP, 453-459.
- Karat, J. (1988) - Software evaluation methodologies. In *Handbook of Human-Computer Interaction*, M. Helander (Ed.), Amsterdam, North-Holland : Elsevier Science Publishers, 891-903.
- Kieras, D.E. et Polson, P.G. (1985) - An approach to the formal analysis of user complexity, *International Journal of Man-Machine studies*, 22, 365-394.
- Moran, T.P. (1981) - The command Language Grammar : A representation of the user interface of interactive computer systems, *International Journal of Man-Machine studies*, 15, 3-50.
- Nielsen, J. et Molich, R. (1990) - Heuristic evaluation of user interfaces. In *Proceedings of the Conference on Human Factors in Computing Systems, CHI'90*, Seattle, April 1-5, 1990, J. C. Chew et J. Whiteside (Eds.), ACM, 249-256.
- Norman, D.A. (1983) - Design principles for human-computer interfaces. In *Proceedings of the Conference of Human Factors in computing Systems, CHI'83*, A. Janda (Ed.), Boston, December 1983, New-York : ACM.
- Pollier, A. (1990) - *Etude descriptive d'une activité de conception-évaluation ergonomique d'un système de messagerie vocale*. Rapport de recherche 1160, Rocquencourt, France : INRIA.
- Reisner, P. (1981) - Formal grammar and human factors design of an interactive

- graphics System, *IEEE Transactions on Software Engineering*, SE.7, (2), March.
- Roberts, T.R. et Moran, T.P. (1983) - The evaluation of text editors : methodology and empirical results, *Communications of the ACM*, 26, 265-283.
- Rosson, M.B., Maass, S. et Kellog, W.A. (1988) - *The Designer as User : Building Requirements for Design Tools from Design Practice*. IBM Research Report, RC 13939, IBM T. J. Watson Research, Yorktown Heights, New York, 1988.
- Savage, R., Habinek, J. et Barnhart, T. (1982) - The design, simulation and evaluation of a menu driven user interface. In *Human Factors in Computer Systems*, Gaithersburg, MD, Mars, 36-40.
- Scapin, D. L. (1986) - *Guide ergonomique de conception des interfaces homme-machine*. Rapport de recherche 77, Rocquencourt, France : INRIA.
- Scapin, D. L., Reynard, P. et Pollier, A. (1988) - *La conception ergonomique d'interfaces: problèmes de méthode*. Rapport de recherche 957, Rocquencourt, France : INRIA.
- Scapin, D. L. (1990a) - Decyphering human factors recommendations. In *Proceedings of the 2nd. International Conference on human Aspects of Advanced Manufacturing and Hybrid Automation*, August 12-16, Honolulu, USA.
- Scapin, D. L. (1990b) - Des critères ergonomiques pour l'évaluation et la conception d'interfaces. In *Proceedings du XXVI<sup>ème</sup> Congrès de la S.E.L.F.* ; Montréal, Octobre 1990.
- Senach, B. (1990) - *Evaluation des interfaces homme-machine : une revue de la littérature*. Rapport de recherche 1180, Rocquencourt, France : INRIA.
- Smith, S. L. et Mosier, J. N. (1984) - *A design evaluation checklist for user-system interface software*. Report MTR-9480 EDS - TR-84-358. The MITRE Corporation, Bedford, MA.
- Streveler, D. J. et Wasserman, A. I. (1985) - Quantitative mesures of the spatial properties of screen design. In *Human computer Interaction, INTERACT'84*, B. Schackel (Ed.), Amsterdam, North-Holland : Elsevier Science Publishers, IFIP, 81-89.
- Sweeney, M. et Dillon, A. (1987) - Methodologies employed in the psychological evaluation of H.C.I.- In *Human-Computer Interaction, INTERACT'87*, H. J. Bullinger et B. Schackel (Eds.), Amsterdam, North Holland : Elsevier Science Publishers, FIP, 367-373.
- Tullis, T. S. (1988) - A system for evaluating screen formats : research and application. In *Advances in Human-Computer Interaction*, H. R. Harston et D. Hix (Eds.), Vol 2 : Norwood, N.J. : 214-286.

## **EXEMPLES D'ETATS DE L'INTERFACE**


## WEC : Etat des Communications (écran d'accueil)






## WC : Configurateur

Gestion des communications		
Configurateur	Emettre appel	Fonctions
<b>Configurateur</b>		
Nombre maximum de communications	5	(5 max)
Communications CUD minitel:	1	(1 max)
Communications CUD TTY :	2	(2 max)
Communication CCBT :	2	(2 max)
Canal préféré en mode paquet		
<input type="radio"/> D uniquement	<input type="button" value="Ok"/>	
<input type="radio"/> B uniquement	<input type="button" value="Quitter"/>	
<input checked="" type="radio"/> B ou D - D prefere		
<input type="radio"/> B ou D - B prefere		
Commutateur	<input checked="" type="radio"/> NUMERIS	<input type="radio"/> CERANE
Sous adresse :	20	

## WNR : Numéroteur RNIS

NUMÉROTEUR RNIS					
Quitter	N° d'Appel	Sous-adresse	Type communication	Message	Répertoire
<b>Caractéristiques du N° demandé</b>					
Numéro d'appel :					
Sous-adresse :					
Type de communication:					
Message :					
					
<b>Actions</b>					
Numéroter					
Raccrocher					
<b>Messages de signalisation</b>					

**RNIS**



## WNA : Numéro d'Appel

**NUMEROTEUR RNIS**

Quitter Sous-adresse Type communication Message Répertoire

**Caractéristiques du N° demandé**

**Numéro d'Appel**

Veillez saisir le N° de téléphone (16 chiffres max)  
ou choisir dans le répertoire

N° de téléphone

N°s téléphone du répertoire




96353100	↑ ↓
96353400	
96353300	
96353000	

Ajouter  
Supprimer  
Quitter  
Ok

**RNIS**

## WSSAD : Sous-adresse

NUMEROTEUR RNIS				
Quitter	N° d'Appel	Type communication	Message	Répertoire
<b>Caractéristiques du N° demandé</b>				
<b>Sous adresse RNIS</b>				
Veuillez saisir la Sous-adresse (4 chiffres max) ou choisir dans le répertoire				
Numéro d	Sous-adresse: 10			
Sous-adr				
Type de				
Message				
<b>Actions</b>		<b>Répertoire des Sous-adresses</b>		
Numér		Ajouter		
Raccro		Supprimer		
		Quitter		
		Ok		

   RNIS

## WSUU : Signalisation usager-usager

NUMEROTEUR . RNIS			
Quitter	N° d'Appel	Sous-adresse	Type communication
Répertoire			

**Signalisation usager\_usager**

Veillez saisir le message(32 caracteres max) ou  
choisir dans le répertoire

Message de signalisation

Répertoire des messages de signalisation

appel téléphonique

appel sur VNI paquet

↑



↓

Ajouter

Supprimer

Quitter

Ok

 RNIS

## WREP : Répertoire


**NUMEROTEUR RNIS**

Quitter N° d'Appel Sous-adresse Type communication Message

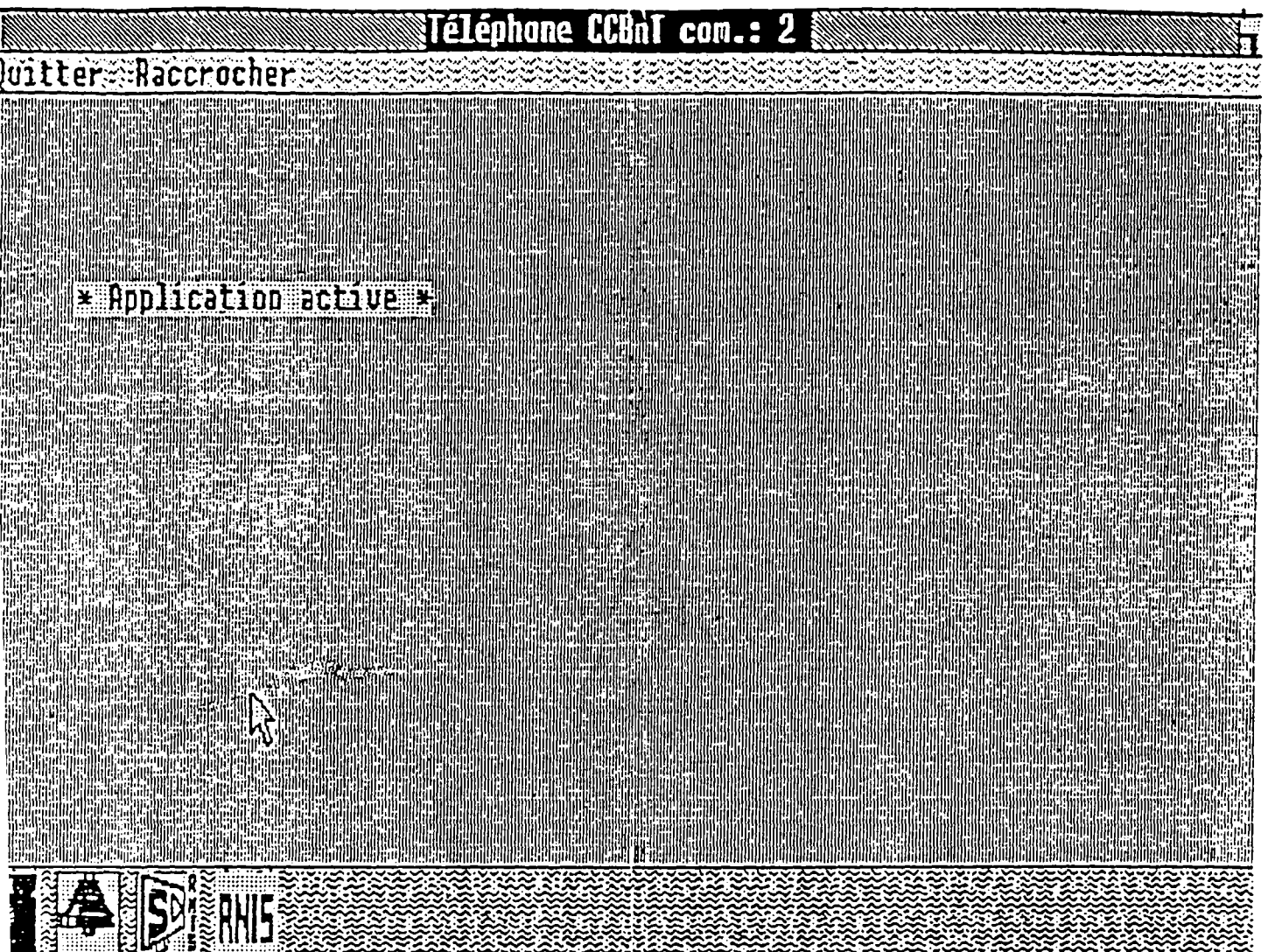
**Répertoire des numéros de téléphone**

Vous pouvez: \* choisir un numéro de téléphone avec ses paramètres.  
\* modifier le répertoire.

Numéro d'Appel	s/ad	Type communication	Message de signalisation
96353300	10	Téléphone CCBnT	RENDEZ-VOUS DU 9/11 REPORTE
96353300	10	Télétype	ENVOI DU DOCUMENT PROMIS L.X.E.
96353000		Téléphone CCBnT	RAPPELER D'URGENCE 96050000
96353200		Téléphone CCBnT	REUNION FIXEE A 14H

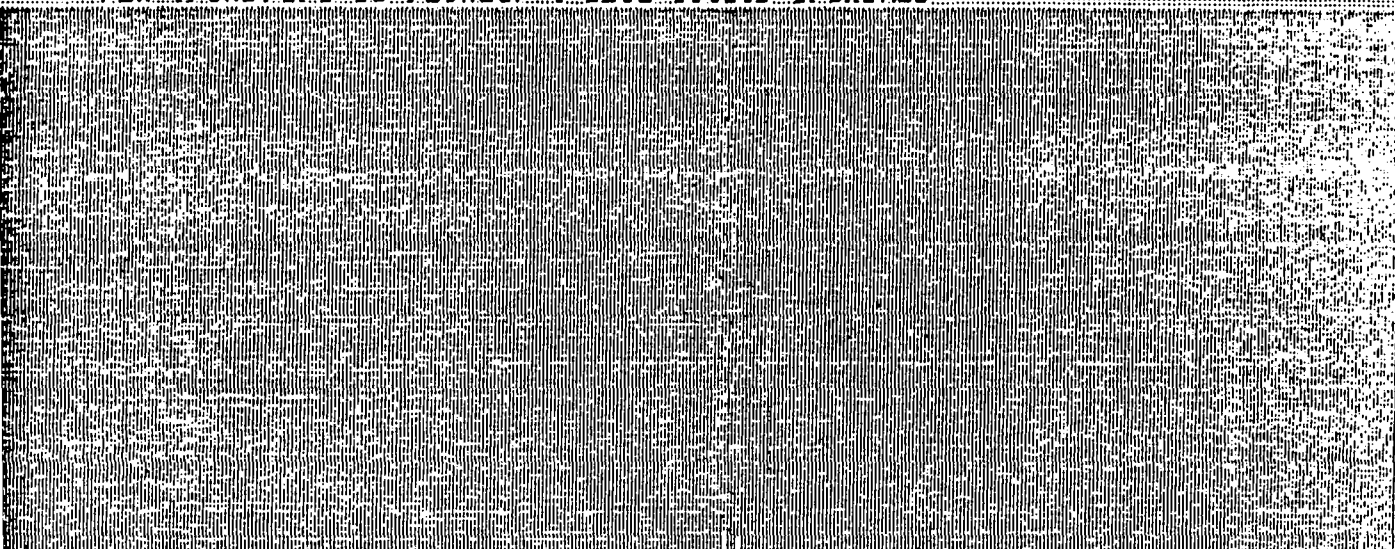

 RNIS

## WAPPACT : Application Active





## WTTF : Télétpe/Transfert de fichier

Télétpe/Transfert Fichier		con.: 1 < Taxes = 2 >	
Quitter Raccrocher Transfert de fichier			
***** Transfert du Fichier : *****			
Nombre octets transmis = 800			
Nombre octets transmis = 1600			
Nombre octets transmis = 2262			
***** Fin Transfert de Fichier : 2262 octets transmis *****			
			
			

## WMESLIB : Message de libération

Téléphone CCBnT con.: 1 < Taxes = 2 >

Quitter Raccrocher

**!! Choix d'un message de libération !!**

Le message à saisir est facultatif

\* Application

arrêt de l'application

arrêt de l'application



message crypté : zen et rn15

Ajouter

Supprimer

Envoi

Quitter

  RNIS

**STRUCTURE TYPOLOGIQUE**  
**ET**  
**EXEMPLE DE CLASSIFICATION DES PROBLEMES EVOQUES**



**STRUCTURE TYPOLOGIQUE****CRITERE****AFFICHAGE****EXISTENCE****Inutile**

En permanence

Dans certains contextes

**Inadéquate**

Position

Position relative

Format (aspect général, taille, forme)

Codage

Entité

Dénomination

Graphique (icône)

Externe

Couleur

Trame (fond)

Soulignement

Autres caractéristiques

Contenu général

**INEXISTENCE****ENTREE & SEQUENCES DE COMMANDE (PROCEDURE)****EXISTENCE****Inutile****Inadéquate****INEXISTENCE**

## CHARGE DE TRAVAIL

### (Actions minimales - Brièveté - Charge mentale)

#### AFFICHAGE

##### EXISTENCE

##### Inutile

##### En permanence

- (Regroupement spatial à tort de l' Icône S avec les autres ; cf 45a ; Guidage) et Icône S affichée en permanence mais inutile et inintéressante pour l'utilisateur

Charge mentale

Icône S (45b)<sup>1</sup>

- Présence d'options qui ne servent à rien pour l'utilisateur dans Menu Fonctions (pour développeurs).

Charge mentale

Menu FONCT (54)

- Présence d'options qui ne servent à rien pour l'utilisateur dans le Configurateur (pour développeurs) ; il faut éliminer tout ce qui ne sert à rien pour l'utilisateur.

Charge mentale

WC (66)

##### Dans certains contextes

- Bouton de commande Raccrocher affiché en permanence mais ne sert à rien si on n'est pas en communication, donc ne doit pas apparaître.

Charge mentale

WNR (39)

- Aire "Message de signalisation" affichée en permanence mais ne sert à rien si on n'est pas en communication, donc ne doit pas apparaître.

Charge mentale

WNR (40)

##### Inadéquate

Position

Position relative

Format (aspect général, taille, forme)

Codage

Entité

Dénomination

Graphique (icône)

Externe

Couleur

Trame (fond)

Soulignement

Autres caractéristiques

Contenu général

##### INEXISTENCE

- (Perte d'informations sur le contexte de la communication dans WAPPACT ; cf 41a et 69a ; Guidage) ; on ne se souvient plus qui appelle ou qui a appelé

Cas de l'appel sortant (41) et Cas de l'appel entrant (69)

Charge mentale - WAPPACT (41b) (69b)

---

<sup>1</sup> a, b fait référence aux critères évoqués simultanément

**CHARGE DE TRAVAIL (suite)****ENTREES & SEQUENCES DE COMMANDE (PROCEDURE)****EXISTENCE****Inutile et/ou Inadéquate = Trop longue<sup>2</sup>**

- Passage inutile par un menu pour remplir un champ ; on devrait pouvoir remplir les champs directement.

Actions minimales

**WNR(10)**

- Pour corriger un champ dans le Numéroteur, obligation d'écraser un champ par un autre ; pas de possibilité de corriger directement un champ.

Actions minimales

**WNR(20)**

- Obligation d'envoyer un message (blanc) pour raccrocher, c'est laborieux, (et puis c'est pas normal ça, je veux raccrocher c'est tout, je ne veux pas de message, ça ne correspond pas à la logique de raccrocher cf 22b ; Compatibilité)

Actions minimales

**WMESS Lib (22a)**

- Pour quitter l'application (c-a-d une communication en cours), obligation de faire 2 fois quitter c'est redondant

Actions minimales

**WAPPACT (23)**

- Passage inutile par une autre fenêtre pour modifier le répertoire ; nécessité de remplir WNR avant, c'est trop compliqué

Actions minimales & Charge mentale

**WREP(29)**

- Option "Com Tel" dans menu "Type de communication" est restée en gras alors que les limitations du système n'autorise pas de la sélectionner lorsqu'il y a déjà une com de ce type en cours; il faut enlever tout ce que à quoi l'utilisateur n'a pas droit, cf 47a ; Guidage), cela lui évite de se lancer dans des procédures qui ne peuvent aboutir.

Actions minimales

**MENU TYCOM (47b)**

- Obligation de repasser par l'appli Gestion des Coms pour sélectionner un autre fichier à transférer ; il faudrait pouvoir choisir directement un autre fichier à partir de cet état.

Actions minimales

**WTTF(52)**

- Pour enchaîner plusieurs appels, procédure trop longue, trop d'actions à enchaîner, obligation de passer par l'appli Gestion des Com entre chaque appel, ce qui est inutile

Actions minimales

**WEC (61)**

<sup>2</sup> Pour le critère "Actions minimales", la distinction inutile et inadéquate n'est pas pertinente pour différencier les problèmes. En effet, inadéquate signifie toujours trop long et on peut aussi interpréter ce type de problème comme la présence d'étapes et/ou d'actions qui sont inutiles.

**CHARGE DE TRAVAIL (suite)**

- Remarque Générale : Compte tenu des fonctionnalités actuellement opérationnelles dans le logiciel alors le dialogue est beaucoup trop compliqué (obligation de faire beaucoup de choses pour faire quelque chose de simple)

Actions minimales

**Général (62)**

- Lorsque l'on rentre dans l'application Gestion des coms passage obligatoire par le configurateur ; étape obligatoire pas nécessaire

Actions minimales

**WC (67)**

- Obligation de passer par fenêtre état des communications ; inutile, pouvoir passer directement à WNR

Actions minimales

**WEC(78) (id61)**

- Obligation de tout remplir WNR pour changer de type de communication une fois la première communication établie (visuo)

Actions minimales

**WNR (80) (id20)**

## SIMILARITE DES PROBLEMES

Tableau 1

Nombres et pourcentages de problèmes évoqués  
selon leur degré de communauté

Sujets	Nombre de problèmes évoqués	% de problèmes évoqués	% problèmes spécifiques vs non spécifiques
S1	30		
S2	30		
S3	17		
S4	21		
<b>Total Spécifique</b>	<b>98</b>	<b>59%</b>	<b>59%</b>
S1S2	7		
S1S3	12		
S1S4	7		
S2S3	1		
S2S4	3		
S3S4	5		
<b>Total Duo</b>	<b>35</b>	<b>21%</b>	
S1S2S3	7		
S1S2S4	5		
S1S3S4	6		<b>41%</b>
S2S3S4	6		
<b>Total Trio</b>	<b>24</b>	<b>14%</b>	
S1S2S3S4	10		
<b>Total Commun</b>	<b>10</b>	<b>6%</b>	
<b>TOTAL</b>	<b>167</b>	<b>100</b>	<b>100</b>

## TYPES DE PROBLEMES EVOQUES

Le tableau 1 présente, par sujet, les pourcentages de problèmes évoqués appartenant aux catégories Affichage et Séquences de commandes. Le pourcentage de problème d'affichage chez chaque sujet correspond au rapport entre le nombre de problèmes d'affichage évoqués chez un sujet et le nombre total de problèmes évoqués par ce sujet (hormis les problèmes fonctionnels). Le nombre de problèmes évoqués appartenant à la catégorie Séquences de commandes a été calculé de la même façon.

Tableau 1

Pourcentages de problèmes évoqués par les sujets, appartenant aux catégories Affichage et Séquences de commandes (par sujet)

Type de problèmes	Sujets			
	Sujet 1	Sujet 2	Sujet 3	Sujet 4
AFFICHAGE	69	77	67	64
SEQ. DE COMMANDES	31	23	33	36

Le tableau 2 présente, par sujet, les pourcentages de problèmes évoqués appartenant aux catégories Existence et Inexistence. Le principe de calcul est le même que celui utilisé pour les catégories Affichage et Séquences de commandes.

Tableau 2

Pourcentages de problèmes évoqués par les sujets, appartenant aux catégories Existence et Inexistence (par sujet)

Type de problèmes	Sujets			
	Sujet 1	Sujet 2	Sujet 3	Sujet 4
EXISTENCE	67	70	82	83
INEXISTENCE	33	30	18	17

pas m'faute !

## CRITERES

Le tableau 1 présente les pourcentages de problèmes évoqués par l'ensemble des sujets, par critère. Le pourcentage de problèmes évoqués pour un critère correspond au rapport entre le nombre de problèmes évoqués par l'ensemble des sujets affectés à ce critère et le nombre total de problèmes évoqués par l'ensemble des sujets affectés aux différents critères (standards et problèmes fonctionnels exclus).

**Tableau 1**  
Pourcentages de problèmes évoqués par l'ensemble des sujets par critère

Critères							
CHTR	COMP	CONS	COUT	ADAP	GEER	GUID	SICO
26	4,5	9	3,5	1	9	28	19

Le tableau 2 présente les pourcentages de problèmes évoqués par chaque sujet, par critère. Le pourcentage de problèmes évoqués par un sujet pour un critère correspond au rapport entre le nombre de problèmes évoqués par ce sujet affecté à ce critère et le nombre total de problèmes évoqués par ce sujet affectés aux différents critères (standards et problèmes fonctionnels exclus).

**Tableau 2**  
Pourcentages de problèmes évoqués par critère et par sujet

Sujets	Critères								Total
	CHTR	COMP	CONS	COUT	ADAP	GEER	GUID	SICO	
Sujet 1	24	3	12	6	4	17	22	12	100
Sujet 2	20	5	6,5	2	0	6,5	40	20	100
Sujet 3	32	5	2	5	0	7	21	28	100
Sujet 4	32	5	11	2	0	2	27	21	100

**Rappel :**

CHTR = Charge de travail ;	COMP = Compatibilité ;
CONS = Consistance ;	COUT = Contrôle Utilisateur ;
ADAP = Adaptabilité ;	GEER = Gestion des Erreurs ;
GUID = Guidage ;	SICO = Signification des Codes.

## STANDARDS ET PROBLEMES FONCTIONNELS

Le tableau 1 présente les pourcentages de problèmes évoqués appartenant à la catégorie Standards, globalement (i.e. pour l'ensemble des sujets) et par sujet. Le pourcentage global correspond au rapport entre le nombre de Standards évoqués par tous les sujets et le nombre total de problèmes ; le pourcentage de Standards évoqués chez chaque sujet correspond au rapport entre le nombre de Standards évoqués chez un sujet et le nombre total de problèmes évoqués par ce sujet.

**Tableau 1**  
Pourcentages de Standards évoqués, globalement et par sujet

	Sujets				
	Tous sujets	Sujet 1	Sujet 2	Sujet 3	Sujet 4
% de Standards	6	9	0	7	9

Le tableau 2 présente les pourcentages de problèmes fonctionnels évoqués, globalement et par sujet. Le principe de calcul est le même que celui utilisé pour les standards ci-dessus.

**Tableau 2**  
Pourcentages de problèmes fonctionnels évoqués, globalement et par sujet

	Sujets				
	Tous sujets	Sujet 1	Sujet 2	Sujet 3	Sujet 4
% de Problèmes Fonctionnels	11	12	8	19	5



*Imprimé en France*  
par  
l'Institut National de Recherche en Informatique et en Automatique

**ISSN 0249 - 6399**